Fordham University
Department of Economics
Discussion Paper Series

Determinants of Schooling Outcomes – Empirical Evidence from Rural Ethiopia

Subha Mani
Fordham University, Department of Economics
John Hoddinott
International Food Policy Research Institute
John Strauss
University of Southern California, Department of Economics

Discussion Paper No: 2009-03
February 2009

Department of Economics
Fordham University
441 E Fordham Rd, Dealy Hall
Bronx, NY 10458
(718) 817-4048

Determinants of Schooling Outcomes - Empirical Evidence from Rural Ethiopia

Subha Mani, Fordham University

John Hoddinott, International Food Policy Research Institute

John Strauss, University of Southern California

January 2009

Acknowledgements:
We thank the following organizations for funding the survey work on which this paper is based: SIDA, USAID and the World Food Programme. We thank John Ham, Jeffrey B Nugent and Jeffrey Wooldridge for comments and suggestions. We are thankful to Yisehac Yohannes for clarifying several data related questions. Financial support was provided by the College of Letters, Arts and Sciences, University of Southern California and International Food Policy Research Institute.

Corresponding author:
Subha Mani**,** Department of Economics, Fordham University, 441 East Fordham Road, Bronx, New York 10458.
Email – smani@fordham.edu

Abstract:

This paper examines the determinants of schooling outcomes - current enrollment status and relative grade attainment - among primary school children in rural Ethiopia. We use repeated cross-sectional data from 15 rural villages in Ethiopia to capture the impact of the changing socioeconomic environment on these outcomes between 1994 and 2004. We find that parental schooling is positively associated with schooling enrollment but its estimated effects declines over time. We observe a similar decline in the estimated impact of father's schooling on relative grade attainment, while the impact of mother's schooling increased during this period. OLS estimates of the impact of household income are biased downwards relative to IV results. Community characteristics are not associated with schooling enrollment. However, the provision of electricity is positively, and distance to primary school negatively, associated with relative grade attainment. These findings suggest that policies that address both supply and demand side constraints have the potential to improve the low levels of schooling attainments found in Ethiopia and elsewhere.

1. Introduction

Sub-Saharan African countries have some of the lowest primary school enrollment rates in the world; only 59% of primary school aged-children were enrolled in school during 1996-2002 (UNICEF country statistics). Primary school completion rates are also low. The World Bank's 2009 World Development Report indicates that only 51% of African children completed primary school. Economic theory predicts and many empirical studies confirm that investments in schooling are positively associated with higher economic and non-economic gains in the future for the individual, her household and the aggregate economy. For example: variants of the Mincer (1974) equation have been widely used to estimate the private returns from schooling. Micro-level findings indicate that in developing countries, the private returns to schooling lie between 5 and 15% (Orazem and King, 2008).].[1] There are large social returns from investing in schooling associated with improvements in maternal and child health and lower fertility rates. Along with existing micro-level benefits, investment in education also contributes towards the economic growth of a country [Mankiw et al., 1992].

The conjunction of high private and social returns to schooling with low enrollments and completion rates points to the value of understanding the determinants of schooling enrollments and grade attainments in Africa. This might appear to be well-trodden territory with studies identifying labor market returns, parental education, household income, and school characteristics as key determinants of schooling outcomes.[2] But closer inspection reveals important limitations to the extant literature. Some studies: (a) are based on samples of individuals who have completed their schooling, these final schooling attainments are explained using socioeconomic characteristics from the individual's current period. Such approaches are potentially misspecified if these characteristics do not map to the year in which the schooling investment decision is made[3]; (b) use data for older children (say those 15 years and older) who have completed schooling. This results in the analysis of a potentially non-random sample of children, since a non-random subset of boys in this age group will have migrated in search of work opportunities and a non-random subset of girls will migrate on marriage; (c) use data on school age children who may not yet have attained their final schooling grade. The majority of the observations in such samples are censored so OLS regressions suffer from censoring bias[4]; (d) rely on single year cross-sectional estimates which are unable to capture the impact of the changing socioeconomic environment on schooling investments across cohorts; and (e) treat household characteristics such as income as pre-determined, thereby ignoring potential correlations between such characteristics and unobserved characteristics that also affect schooling outcomes.

This paper contributes to the literature on the determinants of schooling outcomes in rural Ethiopia taking these methodological issues seriously. First, we estimate the determinants of schooling outcomes including both current enrollment and schooling progression measures (relative grade attainment) of investments in schooling. Current enrollment is a short-run measure of schooling, while schooling progression is a summary measure of schooling from start to the survey. Second, using repeated cross-sectional data allows us to capture the impact of the changing socioeconomic environment on schooling outcomes among primary school children between 1994 and 2004. Third, our analysis sample is restricted to primary school age children

[1] See Psacharopoulos and Patrinos (2004) for the estimated private returns to schooling for over 40 countries.
[2]See Schultz (1988), Rosenzweig (1990), Lillard and Willis (1992), Parish and Willis (1993), Glewwe and Jacoby (1995), Behrman and Knowles (1999) and section 2 below.
[3]See Tansel (1997) and Bommier and Lambert (2000).
[4] See King and Lillard (1983, 1987), Tansel (1998), Holmes (1999), for further discussion.

and the socioeconomic characteristics map to the year in which schooling investment decision is made. Restricting the sample to include primary school age children alone addresses individual specific out migration related selection concerns. Our sample has very little attrition at the household level, addressing potential biases associated with household level attrition. Fourth, we construct a measure for school progression, actual completed grades of schooling divided by potential grades where potential grade is calculated as total number of grades accumulated had the individual completed one grade of schooling by age 7 and continued to accumulate an additional grade of schooling in each subsequent year. The school progression measure constructed here addresses biases arising from right-censoring and allows for delays in grade progression. Fifth, we also use an IV estimation strategy to address the endogeneity problems in our measure of household income.

Ethiopia is an appropriate setting for a study of schooling outcomes. Historically, the level of schooling attainments has been abysmal with more than 50% of the population having never attended school (Demographic Health Survey, 2005). However, the last fifteen years has witnessed a significant increase in primary school enrollment, particularly in rural areas. In 1995, only 15% of primary school children from rural areas were currently enrolled in school (Schaffner, 2004). By 2005, this had more than doubled to 38.8%. Yet little is known about why this change has come about.

In this paper, we estimate reduced form and conditional schooling demand functions that characterize the determinants of schooling outcomes. We use a repeated cross-sectional framework using observations on primary school age children between 7 and 14 years from each of the 1994, 1999 and 2004 waves of the Ethiopian Rural Household Survey (ERHS). Our results indicate that household income is strongly associated with improvements in schooling enrollments. In 1994, household income had no impact in explaining schooling enrollments, whereas by 2004, a 100% increase in household income is associated with a 17 percentage point increase in enrollment probabilities, highlighting the differential impact of income on enrollments among the two cohorts of primary school age children. We also see noticeable income effects associated with relative grade attainment. Parental schooling is positively associated with enrollment but its estimated effects declines over time. We observe a similar decline in the estimated effect of father’s schooling on relative grade attainment, while the impact of mother’s schooling increased during this period. The provision of electricity is positively and distance to primary school is negatively associated with relative grade attainment. Community characteristics are not associated with enrollment.

The paper is organized as follows. Section 2 provides a brief review of the related literature on schooling outcomes. Section 3 describes the data and analyzes summary statistics. Section 4 outlines the theoretical model which guides the empirical specification estimated in section 5. The results are discussed in detail in section 6. Concluding remarks follow in section 7.

2. Literature Review

The determinants of schooling outcomes have been analyzed for a range of outcome variables such as, enrollment status, completed grades of schooling, relative grade attainment, primary school completion, high school completion, grade repetition, test scores, and age at enrollment. The determinants of schooling outcomes can be broadly classified into demand side factors and supply side factors. Demand side factors include measures of parental schooling, household income, household composition, and household's demand for child labor. Supply side factors include school fees, distance to school and school characteristics such as student-teacher ratio, availability of teaching materials and quality of school building.

Among the demand side factors, household income is found to be a key determinant of schooling outcome. Income captures household's access to resources in the long-run and its associated impact on schooling. Behrman and Knowles (1999) report results from their earlier review of 42 studies covering 21 countries and find that, three-fifth's of the schooling indicators used in these studies showed a significant association between household income and schooling. Measures of income and accumulated wealth are associated with - higher enrollment probabilities (Dostie and Jayaraman, 2006), lowering delays in schooling enrollment (Glewwe and Jacoby, 1995), higher levels of schooling completion (King and Lillard, 1987), lowering school withdrawal (Glick and Sahn, 2000) and improved test scores (Brown and Park, 2002). Empirically, it is often difficult to obtain the causal effect of household income on schooling due to the presence of household specific time-invariant unobservables such as household's perceived preferences towards schooling that are likely to bias the estimated coefficient on income. The presence of random measurement error in household income is also likely to bias the OLS estimate on income. Glewwe and Jacoby (1995), Alderman et. al (1996), Tansel (1998), Pal (2004), and Chaudhury et. al (2006), address this endogeneity problem using instrumental variables such as unearned rental income, transfers from abroad, measure of land ownership, characteristics of the head of the household, household demographic composition, and other measures of household assets. Among papers that report both OLS and IV estimation results, the IV estimates on household income are almost always larger than the OLS estimate.

Parental schooling is another important demand side determinant. Parents directly and indirectly affect children's schooling outcomes through time spent teaching at home and indirectly through the choice of schooling inputs. Studies find that measures of parental schooling are positively associated with better schooling outcomes [Brown and Park (2002), Parish and Willis (1993), Singh (1992), Behrman and Wolfe (1984), Alderman et. al (2001), Tansel (1997), Schaffner (2004), Schultz (1988), Strauss and Thomas (1995), Orazem and King (2008)]. Some studies in the literature specifically find that mother's schooling has a greater impact on children's education outcomes compared to father's education [Alderman et. al (2001), Dostie and Jayaraman (2006), Singh (1992)]. Empirically, it is difficult to capture the causal effect of parental schooling outcomes on children's schooling due to the presence of common intergenerational unobservables that affect both children’s and parental schooling outcomes. Mothers' and fathers' schooling outcomes are also influenced by their initial endowments which are likely to be correlated due to non-random matching at the time of marriage. Behrman and Rosenzweig (2002) find that addressing these issues significantly reduces the impact of mother's schooling on children's schooling attainments changing the sign of the coefficient estimate on mother's schooling from positive to negative, while keeping the impact of father's schooling unaffected. Lillard and Willis (1993) explicitly account for the correlation between parent specific unobservables and child specific unobservables and find that almost two-thirds of the impact of parental education on their children's schooling appears to be a direct or indirect consequence of parental schooling, while only one-third can be attributed to unmeasured factors.

Children's opportunity cost of time at both market and non-market work also affects children's schooling outcomes [Rosenzweig and Evenson (1977), Rosenzweig (1990), Pal (2004) and Singh (1992)]. Other determinants of schooling include household composition variables such as number of siblings or children in the household, which have some effect on schooling outcomes, but the sign and significance remain inconclusive [Singh (1992), Parish and Willis (1993)]. Some of the differences in schooling outcomes are also explained through age and sex specific differences in schooling investments.

Supply side determinants of schooling include school characteristics and community infrastructure variables. Improvements in school quality are associated with higher enrollments (Glewwe and Jacoby, 1994) and early/timely enrollments (Bommier and Lambert, 2000). Institutional materials like blackboards also improve schooling outcomes among children. Glewwe and Jacoby (1994) and Glewwe et. al (1995) show that school characteristics like building materials, writing materials and teaching materials are important determinants of test scores and other measures of cognitive development. There are econometric concerns associated with using measures of school characteristics that are directly observable from the child's school. Children with higher ability are likely to be selected into better schools. This selection is likely to bias the coefficient estimates on the school quality variables. Most studies use standard Heckman selection correction methods to account for the choice among different types of school with the aim of reducing the selection bias (Glewwe and Jacoby, 1994), although some recent studies have begun to use experiments (Glewwe et al., 2004). The availability and extent of substitutability between private and public schools also affects the impact of school characteristics on the final demand for schooling (Glick and Sahn, 2006). Among the supply side determinants, community characteristics are likely to be endogenous due to the presence of non-random program placement effects (Rosezweig and Wolpin, 1986). Dostie and Jayaraman (2006) find that the use of community fixed-effects increases the impact of village characteristics on schooling outcomes accounting for purposive program placement effects. [5]

3. Data

The data used in this paper comes from the 1994, 1999, and 2004 waves of the Ethiopian Rural Household Survey (ERHS). The ERHS is a large-scale socioeconomic survey administered in selected rural peasant associations of Ethiopia during 1989-2004.[6] The first wave of the ERHS was fielded in 1989 during which households from 7 farming villages in central and southern Ethiopia were surveyed. In 1989, only a narrow set of questions were administrated as at the time there was no intention of creating a longitudinal data set. In 1994, 6 of the 7 original villages from 1989 (one of the villages could not be re-visited due to civil unrest) and 9 new villages that account for the diverse farming systems practiced in Ethiopia were additionally selected for survey purposes. A total of 15 rural villages were surveyed in 1994 with the aim of constructing a longitudinal data set. In 1994, two waves of the ERHS were administered, the first wave during January-March and the second during August-October. The ERHS subsequently followed households residing in these 15 rural villages during 1995, 1997, 1999 and 2004 [see Dercon and Hoddinott (2004) and Dercon et. al (forthcoming) for more details on survey design]. The survey provides extensive information on household composition, income, consumption expenditure, farm and non-farm assets, ownership and value of land and livestock units, anthropometrics, harvest use and schooling outcomes. In 1997 and 2004, the survey also collected detailed community level information on infrastructure availability, prices of consumption goods, and wage earnings.

In our sample, only one-third of all school age children accumulated some schooling and less than 10% of those completed primary schooling. Hence it is the socioeconomic environment during primary school age that matters the most in determining children's complete trajectory of current and future schooling attainments. Including observations on high school age children would create selection bias, since there occurs huge out migration among high school age females due to

[5] See Glewwe (2002) for detailed review on the supply side determinants of schooling.

[6] The smallest administrative unit in Ethiopia is called a `peasant association', which is sometimes equivalent to one village or a cluster of villages. We use the term ``villages'' and ``peasant association'' interchangeably throughout this paper.

early marriage. Ezra and Kiros (2001) document that 79% of female migrants from Ethiopia migrated at the time of marriage with the average age at the time of marriage being 16 years. Fafchamps and Quisumbing (2005) find that in the ERHS sample, the average age of a bride during first marriage is 17 years. Aside from individual level migration concerns, households are also likely to migrate in search of better schooling opportunities for their children. This can result in a non-random sample of school age children creating attrition bias. The low levels of household attrition rates in the survey allay this concern. In 1994, the ERHS covered 1477 households. In 1999, 1371 of the original 1477 households were re-interviewed and in 2004, 1304 of the original 1477 households were re-interviewed. In 2004, re-contact rate remained as high as 88.2%. Household level attrition is minimal in these rural areas and is supported by other studies using the ERHS.[7] In rural Ethiopia, land is owned by the government and households cannot obtain land if they decide to move to another location. This severely constraints mobility and keeps household level attrition rates low.

Details on enrollment status and completed grades of schooling were first collected in 1994. This feature of the data, together with the need to account for the irregular spacing of the survey rounds means that we only use data collected in the 1994, 1999 and 2004 waves of the ERHS. Mirroring the levels and patterns observed in national statistics, in 1994 only 12.7% of primary school children were enrolled in school. By 2004, this percentage increased by three times to 45.49%. Four additional observations emerge from descriptive statistics presented in figures 1 and 2. First, the significantly large improvements in enrollments during 1994-1999 reflect mostly new enrollments whereas improvements in enrollments during 1999-2004 reflect both new and continued enrollments. Second, fewer children are enrolled at young ages with enrollments peaking after age 11 depicting delays in average enrollments. Third, there exists a strong positive association between male enrollments and age. However, for females the relationship between age and enrollment is non-linear with enrollments peaking around age 12. Lastly, much of the improvement in enrollment between 1994 and 1999 is concentrated among male children, magnifying male-female differences in schooling attainments. This gender gap narrows during 1999-2004 reducing the overall gender differences in schooling outcomes. Figures 3 and 4 depict trends in relative grade attainment among primary school age children. Timely enrollments in 1999 led to a steep increase in the relative grades accumulated. The pattern of improvement in relative grades is similar to the pattern observed for enrollment rates. Male-female differences in relative grades increase between during 1994-1999 and then narrow during 1999-2004.

4. Model

We outline a model of the determinants of schooling as a means of guiding the selection of variables that appear as regressors in the subsequent sections.

Assume that households maximize utility, U (1), subject to an income constraint (2), and a schooling production function (3).

$$U = u(C_t, S_t, L_t; D_t) \qquad (1)$$

$$p_t^c C_t + p_t^m M_t = w_t(T_t - L_t) + \pi_t \qquad (2)$$

$$S_t = f(M_t, I_t; \theta_c, \theta_{ct}, \mu_h, \mu_{ht}) \qquad (3)$$

[7] Dercon et. al (2006) report household level attrition between 1994 and 2004 to be around 12.4% which is very close to the numbers reported in this paper.

The utility function depends upon food and non-food consumption goods, C_t, leisure, L_t, child's schooling outcome, S_t. It is assumed that the household does not derive any direct utility from the consumption of M_t except via its impact in determining S_t. S_t is modeled here as a pure consumption good from which the household derives utility.[8]

P^c_t is a vector of price of food and non-food consumption goods, P^m_t is a vector of price of schooling inputs, and w_t, is wage rate. T_t, is parents total time endowment. Profit income from farm and non-farm activities and all other sources of non-labor income is captured by π_t. The utility function is affected by time-varying household demographic characteristics D_t such as mother's age and age of the head of the household capturing household experience and life-cycle position.

The schooling production function specified (3) follows the health production function specified in Sahn and Fedorov (2005); Strauss and Thomas (1995, 2008); Mani (2007). S_t is written as a function of schooling inputs, community resources, child characteristics and household characteristics. Schooling outcomes, S_t include measures such as enrollment status, grade attainments, and test scores. Schooling inputs M_t include books, school uniform, and other home inputs. Environmental characteristics, I_t capture overall resource availability in the community and include factors such as number of schools, access to electricity and other community infrastructure that affects schooling outcomes. θ_c, θ_{ct} include child specific characteristics such as child's sex and age capturing age and gender specific differences in the accumulation of schooling outcomes. θ_c, θ_{ct} also include time-varying and time-invariant measures of child's own innate ability to perform well in school capturing overall cognitive development and learning potential. μ_h, μ_{ht} capture household demographic characteristics and other time-invariant and time-varying rearing and caring practices, all of which affect schooling attainments.

Using simple first-order conditions, we can solve for the optimal amount of schooling input, M_t^* as follows:

$$M_t^* = f(p_t^c,\ p_t^m,\ w_t, \mathrm{I}_t,\ \pi_t, \mathrm{D}_t,\ \mu_{ht}, \mu_h, \theta_{ct}, \theta_c) \qquad (4)$$

The static conditional schooling demand function (5) can be obtained by replacing M_t in equation (3) by M_t^* from equation (4):

$$S_t^* = f(p_t^c,\ p_t^m,\ w_t, \mathrm{I}_t,\ \pi_t, \mathrm{D}_t,\ \mu_{ht}, \mu_h, \theta_{ct}, \theta_c) \qquad (5)$$

[8] Some studies specify the utility maximization problem as one in which the household derives utility only from the expected future returns from investing in children's schooling [Rosenzweig (1990), Bommier and Lambert (2000), Brown and Park (2002)]. This requires a reliable measure on wage earnings, which is not available to us for all waves of the ERHS. A related approach assumes that parents seek to equalize the expected future returns from schooling across all their children. [Behrman, Pollak and Taubman (1982)]. Implementing this approach requires data on child ability as well as earnings neither of which is available to us.

5. Empirical Specification

The empirical counterpart of the conditional schooling demand function can be written as follows.

$$S_i = \beta_0 + \sum_{j=1}^{R} \beta_j X_{ij} + \varepsilon_i + \varepsilon_h + \varepsilon_v \qquad (6)$$

where S_i is the enrollment status and relative grade attainment of child i. Enrollment status is defined as a dummy variable equaling one if the child is enrolled in school at the time of the survey, zero otherwise.[9] Relative grade attainment is defined as actual grades divided by potential grades where potential grades is calculated as total number of grades accumulated had the individual completed one grade of schooling by age 7 and continued to accumulate an additional grade of schooling in each subsequent year.

The Xs include individual, household, and village level variables. At the individual level, we control for the age of the child, male dummy, mother's age and measure of parental schooling. Age is specified using dummy variables with a separate dummy variable assigned for each year between 7 and 14 years, the omitted category is 7 years. The age dummies capture for age specific differences in schooling attainments. The male dummy equals one if male, 0 if female. It captures gender specific differences in schooling outcomes. We interact the age dummies with the male dummy to capture age-gender specific differences in schooling attainments. Parental grade attainment is low, averaging only one grade in 1994 and two grades by 2004. The majority of parents have no formal schooling. For this reason, we characterize parental schooling using dummy variables, equaling one if the child's mother (father) has at least one grade of formal schooling, zero otherwise. Mother's age is included in the regressions to capture mother's experience and knowledge which affects her ability in making household decisions and schooling related decisions.

Household level regressors include number of adult (>18 years) males and number of adult (>18 years) females capturing household demographic composition. Age of the head of the household is included as an additional regressor to capture household experience and life-cycle position. Our empirical specification must also include a measure for household income. We use household consumption expenditure as our measure of long-run income, since it is less likely to be measured with error and, where households smooth consumption over time, is a better representation of permanent income than current income [Thomas, Strauss and Henriques (1990) [10], Behrman and Knowles (1999)]. Total household consumption expenditure is constructed as the sum of value of food items (questionnaire included details on 33 specific food items) consumed including purchased and non-purchased consumption goods (consumption out of own stock), and value of non-investment type non-food items purchased. Non-food items include consumables such as matches, batteries, kerosene but exclude expenditure on durables such as housing [Dercon, et. al forthcoming]. Consumption is valued using prices obtained from market survey fielded at the same time as the household surveys. Total household consumption expenditure is divided by household size to capture the per person resource availability in the household. The nominal per

[9] Some children in our sample are enrolled in religious schools. Our interest is limited to measuring human capital accumulated through learning subjects like mathematics, science and social science; none of which is taught in religious schools. For this reason, we treat children enrolled in religious schools as not enrolled.

[10] Thomas, Strauss and Henriques (1990) were the first to use household per capita consumption expenditure as a measure of long-run income in a paper assessing the determinants of child health.

capita consumption values are then converted to real per capita consumption expenditure using a food price index. We use the logarithm of the real per capita household consumption expenditure to capture non-linearities in the relationship between this characteristic and schooling outcomes. In our model, log PCE is endogenous – it reflects for example labour supply decisions – and so we treat it as endogenous, see below.

The village level regressors are included only for the 2004 regressions since information on village characteristics is only available in 2004. We control for price of schooling using a measure on distance to primary school in km. We also include two measures of community infrastructure. Our first measure captures availability of electricity in the community, it is measured as a dummy variable which takes a value 1 if electricity is available and 0 otherwise. Our second measure captures availability of piped water in the community, it is measured as a dummy variable which take a value 1 if piped water is available and 0 otherwise.

There are three sources of unobservables in equation (6), ε_i, ε_h, and ε_v. ε_v captures village specific unobservables such as village endowments. ε_h captures household specific unobservables such as parental preferences and time-discount rate. We cannot use household fixed-effects to remove ε_h since we do not have enough instances of multiple children from the same household[11]. ε_i is assumed to be a random i.i.d. error term.

Table 1 provides sample averages and standard deviation for all schooling outcome variables and regressors used in the regression specifications.

6. Results

6.1 Enrollment

A linear probability model (LP) is specified to characterize the determinants of schooling enrollment. We use village level fixed-effects with the LP model to obtain unbiased parameter estimates in the face of possible correlation with ε_v. LP allows easy interpretation of the parameter estimates in terms of probabilities and the village level fixed-effects removes all sources of common village level unobservables. A further attraction of the inclusion of these fixed effects is that it addresses cluster-related issues in the standard errors since common village level unobservables are also cluster effects.[12] Further, our estimates account for arbitrary forms of heteroskedasticity using the White (1980) formulation [see Wooldridge (2002)].

We test whether the three survey rounds should be pooled. We obtain an F-statistic of 4.33 which, at the 1% significance level, rejects the null of pooling across rounds. Hence, schooling enrollment regressions are estimated separately for the each survey year. These are reported in appendix tables A1, A2 and A3.[13] Our preferred estimates are found in column 3. Age dummies, male dummy, and age interacted male dummies are included to account for age, gender, and age-

[11] In our 1994 sample, we have 385 households with one primary school age child. In 1999, we have 388 households with one primary school age child and in 2004; we have 354 households with one primary school age child.

[12] We thank Jeffrey Wooldridge for his assistance on this point.

[13] We also include for dummy variables to capture missing observations in our socioeconomic characteristics (if any) that were replaced by the sample mean in all our regression results. The imputations are done for missing observations in parental schooling, mother's age and age of the head of the household. On an average, less than 2% of observations are missing for age of the head of the household. For mother's age, the average is around 10%. For parental schooling variables, the missing percentage is slightly higher 15%.

gender specific differences in enrollments.[14] The estimated coefficients indicate a strong positive association between age and enrollment. The coefficient estimates on age dummies from column 3 table A1 imply that the probability of a 13 year old being enrolled is 11 percentage points higher than the probability of a 11 year being enrolled. The parameter estimates from the 1999 regressions depict increase in enrollment probabilities among all age groups capturing both timely enrollments and continued enrollments. By 2004, the probability of an 8 year old being enrolled is statistically significant (insignificant in earlier rounds) and is 8 percentage points higher than the probability of an 8 year old being enrolled in 1994, capturing improvements in timely enrollments. While these results reflect significant delays in enrollments, they also show these declining over time.

The coefficient estimates on the parental schooling variables reported in column 3 of appendix tables A1-A3 indicate a strong positive relationship between parental schooling and enrollment status. A child whose mother has non-zero grades of schooling is 9 percentage points more likely to be enrolled in 1994, 4 percentage points more likely to be enrolled in 1999, and 6 percentage points more likely to be enrolled in 2004 compared to a child whose mother has no schooling. A child whose father has any schooling is 10 percentage points more likely to be enrolled in 1994, 7 percentage points more likely to be enrolled in 1999 and 6 percentage points more likely to be enrolled in 2004 compared to a child whose father has zero accumulated grades of schooling. A limitation of these results is that we treat parental schooling variables as exogenous. That said, unobserved heterogeneity is likely to be of less concern here since our measure of parental schooling does not distinguish between parents with 10 grades of schooling and 2 grades of schooling. It is differences in actual completed grades of schooling that is more likely to be confounded by differences in innate abilities compared to the qualitative measure of parental schooling used here.

Children of educated parents may be more likely to be enrolled on time (at an earlier age) compared to a child of a less educated parent. To assess this possibility, we added interaction terms between the age dummies and the parental schooling variables as additional regressors in our preferred specification. These interaction terms were all jointly insignificant at 1% significance level for all the three enrollment regressions.[15]

If schooling is a normal good, an increase in household permanent income should lead to higher investments in schooling. We use the logarithm of real per capita household consumption expenditure (PCE) as our measure of permanent income. However, OLS estimates of PCE are likely to be biased and inconsistent due to - (1) the potential correlation between household specific time-invariant unobservables (parent's preferences and time discount rate) and PCE, resulting in omitted variables bias, and (2) the presence of random measurement error in data

---

[14] We test if the socioeconomic characteristics controlled in the regression specifications vary by gender. A joint test on the interaction between the gender dummy and the socioeconomic characteristics from 1994, yields an F statistic of 1.11 (p-value =0.35). A similar test on the pooled sample from 1999 yields an F statistic of 1.39 (p-value =0.20) and from 2004 yields an F statistic of 0.62 (p-value =0.80). We conclude that the impact of these socioeconomic characteristics do not vary by gender and hence, we estimate our static model pooling boys and girls. The age and gender interaction terms in the pooled specifications allow for age-gender specific differences in schooling attainments.

[15] The chi2 statistic on the interaction between the age dummies and mother's schooling for the enrollment regressions in 1994, 1999 and 2004 are - 4.78 (0.68), 10.28 (0.17) and 9.10 (0.24) respectively with p-values in parenthesis. The chi2 statistic on the interaction between the age dummies and father's schooling in the enrollment regressions from 1994, 1999 and 2004 are - 10.61 (0.15), 6.93 (0.43) and 5.97 (0.54) respectively with p-values in parenthesis.

which biases the estimated coefficient on PCE towards zero. Given these concerns, we use instrument variables that provide unbiased and consistent estimates on household income.

We use land, livestock units and an interaction term between land and rainfall as excluded instruments for PCE in our first-stage regressions. [16] The ERHS provides details on land holdings and livestock units, two forms of assets that households own in rural Ethiopia. In a static framework, all forms of assets (land and livestock units) can be treated as being exogenously determined; they are treated as given to the household. In Ethiopia, land allocations are based on local administrative decisions; it is illegal to buy or sell land. These allocation are determined independent of household's schooling investment decision and hence can be treated as exogenous to the household. Households in this sample rely heavily on income from agriculture which is rainfall dependent (Dercon, Hoddinott and Woldehanna, 2008; Dercon et al, forthcoming.). The extent to which rainfall affects household income will also depend upon the amount of land controlled by the households; hence, we interact land with rainfall. Rainfall is measured in terms of the average amount rainfall (in mm) during the main cropping season in each village. Conceptually we argue that land, livestock units and interaction between land and rainfall are valid instruments for PCE reported in column 3 of tables A1-A6.

The preferred IV estimates of log (PCE) are reported in column 3 of tables A1-A3, a summary of which is reported in table 2. Also included are the IV estimates on log (PCE) as reported in column 4 of table A3 where the village fixed-effects are replaced with the actual village level characteristics.

The preferred IV estimate on PCE is almost three times larger than the OLS estimate. The OLS estimates reported in table 2 indicates that a 100% increase in household income increases enrollment probability by a tiny amount, 3 percentage points in 1994 and is slightly larger in 2004, 6 percentage points. While the IV estimates for 1994 show no effect on enrollment, and the IV estimates for 2004 indicate a much larger magnitude; a 100% increase in household income increases the probability of being enrolled by 17 percentage points. The large differences between the OLS and IV estimates capture the magnitude of biasedness in the OLS parameter estimate. The increase in the coefficient estimate from OLS to IV indicates that the OLS estimate of PCE is likely to be biased downward due to measurement error and not biased upwards due to omitted variables. Some papers in the literature have not accounted for the endogeneity in the household income variable and hence their estimated coefficient on household income is likely to be biased and inconsistent [Bommier and Lambert (2000), Dostie and Jayaraman (2006)].We also find that these income effects increased most between 1994 and1999 and reduce a little between 1999 and 2004. These changes are similar to the average changes in household income observed during this period. There was a 25% increase in average household income between 1994 and 1999 and almost no change between 1999 and 2004.

Our preferred IV estimates are robust to concerns regarding instrument validity. Instruments are considered valid only if they are strongly correlated with the endogenous regressor and uncorrelated with the error term in the second-stage regressions. The instruments used here are both strongly correlated with the endogenous regressor and uncorrelated with the error term in the second-stage regression. Staiger and Stock (1997) suggest that in the presence of a single endogenous regressor, instruments are deemed to be weak if the first-stage F statistic on the excluded instruments is less than 10. Following this rule, we find that the F statistic on the excluded instruments reported in our regressions are well over 10 almost always, rejecting the null of weak correlation between the instruments and the endogenous regressor. An alternate test

[16]The first-stage regression results are provided in appendix table A7.

of weak instruments given by the Kleibergen-Paap Wald rk F statistic is robust to the presence of heteroskedasticity, autocorrelation and clustering [Kleibergen and Paap, 2006]. However, in the presence of a single endogenous regressor, the Kleibergen-Paap test statistic reduces to the First-stage F statistic on the excluded instruments as reported here. The Hansen J statistic is also appended in our regression tables, and we find that we cannot reject the null that the instruments are uncorrelated with the error term and appropriately excluded from the second-stage regressions. Note too that our results are robust to the choice of instruments. We test this by replacing current period livestock units with lagged livestock units and by replacing rainfall during the main cropping season with lagged rainfall. Parameter estimates obtained on income are not statistically different between these alternate IV specifications.

Section 2 noted the empirical literature on the role of the supply side determinants of schooling outcomes. Our 2004 survey round collected information that allows us to explicitly control for these. We drop the village dummy variables and include distance to the nearest primary school (measured in km) and dummy variables equaling one of the village has access to electricity and if it has piped water. The coefficient estimates on these supply side characteristics are correctly signed but are not statistically significant. The coefficient estimates on the demand factors are unchanged.

6.2 Relative Grade Attainment

Schooling enrollments depict short-term investments in education. They do not capture regular attendance and grade advancement. To capture long-term investments in schooling and grade advancement, we use completed grades of schooling as an outcome variable of interest. However, because many children in the sample have not yet started schooling, a large number of observations are censored at zero. Further, observations on completed grades of schooling will be right-censored for children currently enrolled in school. Both sources of censoring make standard OLS estimates on the right hand side variables biased and inconsistent (Tansel, 1997).

One way to address this problem is to restrict the sample to include observations on children with completed schooling spells alone. Such a sample can be estimated using an ordered probit specification to obtain unbiased and consistent parameter estimates. However, this is likely to create out migration related selection concerns as it includes observations on high school age children and older populations. Also the right hand side variables used to characterize the determinants of completed grades may not be representative of the actual socioeconomic environment that affected final schooling attainment. By contrast, King and Lillard (1983, 1987) use a censored ordered probit specification. They use maximum likelihood framework in which children who have completed their entire schooling spells (uncensored observations) and children who have not completed their entire schooling spell (censored observations) enter the likelihood function separately. This addresses both sources of censoring bias. However, it relies on the strong assumption that children who belong to the uncensored category do not re-enter schools. Our sample includes observations on primary school age children, who are likely to re-enter school at a much later date and or start schooling at a much later date since they have never been enrolled, making this estimation strategy infeasible in this context.

An alternative approach involves creating a relative measure of completed grades of schooling. Birdsall (1982) defines schooling as actual grades divided by the mean grades for the relevant age-sex category. The advantage of using relative measures of schooling is that a continuous outcome variable is created making OLS estimates consistent. The relative measure also accounts for delays in enrollments and grade attainments. Individuals with the same completed grades of schooling are treated differently depending upon their age, except if the actual completed grade is zero. The OLS estimates of relative grade attainment may suffer from censoring bias, especially for those individual's who may start schooling at later time and accumulate more than 0 grades of schooling. This downwardly biases the estimated coefficients.

We use the same control variables in the relative grade attainment (RGA) regressions as we used in the enrollment regressions. Our preferred IV estimates with village-level fixed-effects are reported in columns 3 of appendix tables A4, A5 and A6.[17]

In 1994, only 13% of primary school age children were enrolled in school and even fewer had non-zero grades of schooling. There is little variation in relative grades by age and hence no significant relationship between age and relative grades in 1994. By 2004, relative grades systematically improved among all ages and some age dummies are statistically significant. There is no evidence for age specific gender differences in relative grades attainment.

The impact of father's schooling on relative grade attainment is highest in 1994 but declines by 2004, see column 3 of tables A4-A6. While the impact of mother's schooling increases between 1994 and 2004. During 1994-2004, the percentage of women with non-zero completed grades of schooling increased contributing towards the increase in schooling enrollments and hence indirectly causing improvements in relative grade attainments too. The role played by father's schooling consistently falls as the relative importance of income increases in determining schooling attainments. Household income and fathers' schooling can potentially be correlated and this correlation could reduce the estimated impact of father's schooling on relative grade attainment. To capture potential non-linear effects of parental schooling, we interacted the age dummies and the parental schooling variables and added these to our preferred specification. These were jointly insignificant at the 1% significance level in all years.[18]

The preferred IV estimates of log (PCE) are summarized in table 3. The OLS estimates show that a 100% increase in income increases relative grade attainments by 0.02 in 1994 and 0.04 in 2004. The IV estimates show that a 100% increase in PCE increases relative grades by 0.10 in 1994 and is unchanged in 2004. As income increases, the gap between actual grades and potential grades decreases, improving grade progression. The IV estimates on household income are greater than the OLS estimates indicating the presence of omitted variables problem in the OLS regressions, similar to the enrollment results.

Distance to primary school in km and the availability of electricity in the village, both have a statistically significant impact on relative grade attainment (column 3, table A6) while distance to school has a negative impact. Provision of electricity has positive effect capturing community infrastructure and its effect on relative grades. The village supply side characteristics are all jointly significant at 1%.

[17]A pooling test combining the sample from 1994, 1999, and 2004 waves yields an F statistic of 4.33, statistically significant at the 1% level, indicating that we can reject the null of pooling all three waves together. A joint test on the interaction between the gender dummy and the socioeconomic characteristics from 1994, yields an F statistic of 1.11 (p-value=0.35). A similar test on the pooled sample from 1999 yields an F statistic of 0.65 (p-value=0.71) and from 2004 yields an F statistic of 1.38 (p-value=0.18). We conclude that the impact of the socioeconomic characteristics included in these regressions do not vary by gender. Our pooled specifications allow for age-gender specific differences in schooling.

[18]The chi2 statistic on the interaction between the age dummies and mother's schooling for the relative grade attainment regressions in 1994, 1999 and 2004 are - 2.51 (0.92), 8.73 (0.27) and 7.13 (0.41) respectively with p-values in parenthesis. The chi2 statistic on the interaction between the age dummies and father's schooling in the relative grade attainment regressions from 1994, 1999 and 2004 are - 6.65 (0.46), 14.34 (0.20) and 6.95 (0.43) respectively with p-values in the parenthesis.

These results are consistent with much of the literature assessing the determinants of schooling as discussed in the literature review section. There are two papers that are closely related to our work, as they assess the determinants of schooling outcomes among children from Ethiopia, including rural areas in their samples. Chaudhury et. al (2006) estimate the impact of weather induced income shocks to explain the gender differences in schooling outcomes. They pool their samples on primary school children from 1996 and 2000, finding that rainfall shocks are negatively associated with enrollment probabilities and that household income and schooling of the head of the household are all positively associated with higher enrollments. Most of the households in rural Ethiopia rely on agricultural output for income, which is often strongly associated with rainfall. The measure of household income and rainfall used in the second-stage regressions are thus potentially correlated with each other. They also find distance to primary school to be negatively associated with schooling enrollments. Schaffner (2004) uses a repeated cross-sectional data to assess the determinants of schooling outcomes among primary school children. She finds household income and distance to primary school as two key determinants of schooling enrollments outcome.

In addition to the contributions outlined in the introduction section, we also compare our work with these two papers that use data from Ethiopia. Both these studies use short-run measures of schooling for primary school children, while our study examines the determinants of schooling outcomes among primary school children using both short-run and long-run measures of schooling investment - enrollment status and relative grade attainment. Our paper uses repeated cross-sectional data covering changes in socioeconomic characteristics and their effects on primary school outcomes among different cohorts of children for an entire decade between 1994 and 2004. Unlike earlier work, our paper establishes a theoretical framework to guide the empirical specification and relies on the use of completely exogenous instruments that are determined outside the schooling investment decision.

7. Conclusion

In this paper, we examine the determinants of schooling outcomes in rural Ethiopia. We consider both current enrollment status and relative grade attainment and in so doing can consider the determinants of schooling demand in the short run (enrollment) as well as a summary measure of completed schooling (relative grade attainment). We use repeated cross-sectional data to capture the impact of the changing socioeconomic environment on schooling outcomes among primary school children between 1994 and 2004. Our analysis is restricted to primary school age children and the socioeconomic characteristics we consider are mapped to the year in which schooling investment decisions are made. Restricting the sample to include primary school age children alone addresses individual specific out migration related selection concerns. Our measure of grade attainment addresses biases arising from right-censoring and allows for delays in grade progression. We use an IV estimation strategy to address the endogeneity problems associated with the inclusion of household income in our regressions. In so doing, our results are not subject to the methodological flaws that plague much of the literature on the determinants of schooling in Africa and elsewhere.

Parental schooling is positively associated with enrollment but its estimated effects declines over time. We observe a similar decline in the estimated effect of father's schooling on relative grade attainment, while the impact of mother's schooling continued to increase during this period. OLS estimates of the impact of household income are biased downwards relative to IV results. Further, we find an increase in the impact of income on enrollment and relative grade attainment between 1994 and 1999, a period in which household incomes grew markedly. Community characteristics are not associated with enrollment. However, the provision of electricity is positively and distance to primary school is negatively associated with relative grade attainment. These findings suggest that policies that address both supply and demand side constraints have the potential to improve the low levels of schooling attainments found in Ethiopia and elsewhere in Africa.

References

Alderman, H., J.R. Behrman, D.R. Ross and R. Sabot. 1996. Decomposing the Gender Gap in Cognitive Skills in a Poor Rural Economy. The Journal of Human Resources, 31(1):229-254.

Alderman, H., J.R. Behrman, and R. Menon. 2001. Child Health and School Enrollment: A longitudinal analysis. The Journal of Human Resources, 56: 185-205.

Behrman, J.R. and B.L. Wolfe. 1984. The Socioeconomic Impact of Schooling in a Developing Economy. Review of Economics and Statistics, 66(2): 296-303.

Behrman, J.R. and J.C. Knowles. 1999. Household Income and Child Schooling in Vietnam. The World Bank Economic Review, 13(2): 211-256.

Behrman, J.R. and M.R. Rosenzweig. 2002. Does Increasing Women's Schooling Raise the Schooling of the Next Generation? American Economic Review, 92: 323-334.

Behrman, J.R., R.A. Pollak and P. Taubman. 1982. Parental Preferences and Provision for Progeny. The Journal of Political Economy, 90(1): 52-73.

Birdsall, N. 1982. Child Schooling and the Measurement of Living Standards. LSMS working paper

Birdsall, N. 1985. Public Inputs and Child Schooling in Brazil. Journal of Development Economics, 18 (1): 67-86.

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

Brown, P.H. and A. Park. 2002. Education and Poverty in Rural China. Economics of Education Review, 21: 523-41.

Chaudhury, N., L. Christiansen and M. Asadullah. 2006. School, Household Risks and Gender: Determinants of Child Schooling in Ethiopia, working paper, World Bank.

Dercon, S. and J. Hoddinott. 2004. The Ethiopian Rural Household Survey: Introduction. mimeo, Washington D.C: International Food Policy Research Institute.

Dercon, S., J. Hoddinott and T. Woldehanna. “Growth, poverty and chronic poverty in rural Ethiopia: Evidence from 15 Communities 1994-2004,” mimeo, Washington D.C: International Food Policy Research Institute.

Dercon, S., D.O. Gilligan, J. Hoddinott and T. Woldehanna. 2006. The Impact of Roads and Agricultural Extension on Consumption Growth and Poverty in Fifteen Ethiopian Villages, American Journal of Agricultural Economics, forthcoming.

Dostie, B. and R. Jayaraman. 2006. Determinants of School Enrollment in Indian villages. Economic Development and Cultural Change, 54: 405-421.

Ezra, M. and G.E. Kiros. 2001. Rural Out-Migration in the Drought Prone Areas of Ethiopia: A Multilevel Analysis. International Migration Review, 35(3): 749-771.

Fafchamps, M. and A. Quisumbing. 2005. Marriage, Bequest and Assertive Matching in Rural Ethiopia. Economic Development and Cultural Change, 53: 347-380.

Glewwe, P. and H. Jacoby. 1994. Student Achievement and Schooling Choice in Low-income Countries: Evidence from Ghana. Journal of Human Resources, 29(3): 843-864.

Glewwe, P. and H. Jacoby. 1995. An Economic Analysis of Delayed Primary School Enrollment in a Low Income Country: The Role of Early Childhood Nutrition. The Review of Economics and Statistics, 77: 156-69.

Glewwe, P. 2002. Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes. Journal of Economic Literature, XI: 463-82.

Glewwe, P., M. Kremer and S. Moulin. 2004. Retrospective vs. Prospective Analyses of School Inputs: The Case of Flipcharts in Kenya, Journal of Development Economics, 74:251-268.

Glick, P. and D. Sahn. 2000. Schooling of Girls and Boys in a West African Country: The Effects of Parental Education, Income, and Household Structure. Economics of Education Review, 19(1):63-87.

Glick, P. and D. Sahn. 2006. The Demand for Primary Schooling in Madagascar. Journal of Development Economics, 79(1):118-145.

Holmes, J. 1999. Measuring the Determinants of School Completion in Pakistan: Analysis of Censoring and Selection Bias, working papers, Economic Growth Center, Yale University.

King, E.M. and L.A. Lillard. 1983. Determinants of Schooling Attainment and Enrollment Rates in the Philippines. A Rand Note, Santa Monica: Rand Corporation.

King, E.M. and L.A. Lillard. 1987. Education Policy and Schooling Attainment in Malaysia and the Philippines. Economics of Education Review, 6(2):167-181.

Kleibergen, F.and R. Paap. 2006. Generalized reduced rank tests using the singular value decomposition. Journal of Econometrics ,133 (1): 97-126.

Lavy, V. 1996. School Supply Constraints and Children's Educational Outcomes in Rural Ghana. Journal of Development Economics, 51: 291-314.

Lillard, L. and R.J. Willis. 1994. Intergenerational Educational Mobility: Effects of Family and State in Malaysia. The Journal of Human Resources, 29(4): 1126-66.

Maddala, G. S. 1983. Limited-dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press.

Mankiw, G.N., D. Romer, D.N. Weil. 1992. A Contribution to the Empirics of Economic Growth. The Quarterly Journal of Economics, 107(2): 407-437.

Mani, S. 2007. Role of the Household and the Community in Determining Child Health. WIDER Research paper No. 2007/X, Helsinki: UNU-WIDER.

Mincer, J. 1974. Schooling, Experience, and Earnings. New York: National Bureau of Economic Research.

Orazem, P.F. and E.M. King. 2008. Schooling in Developing Countries: The Roles of Supply, Demand and Government Policy. Handbook of Development Economics, Volume 4, edited by T. P. Schultz and J. Strauss.

Pal, S. 2004. How Much of the Gender Difference in Child School Enrollment can be Explained? Evidence from Rural India. Bulletin of Economic Research, 56(2): 0307-3378.

Parish, W.L. and R.J. Willis. 1993. Daughters, Education, and Family Budgets: Taiwan Experiences. Journal of Human Resources, 28(4): 863-98.

Psacharopoulos G. and H.A. Patrinos. 2004. Returns to Investment in Education: A further Update. Education Economics, 12(2): 111-134.

Rosenzweig, M.R. 1990. Population Growth and Human Capital Investments: Theory and Evidence. Journal of Political Economy, 98:S38-S70.

Rosenzweig, M.R. and R.E. Evenson. 1977. Fertility, Schooling and the Economic Contribution of Children in Rural India. Econometrica, 45:1065-1079.

Rosenzweig, M.R. and K.J. Wolpin. 1986. Evaluating the Effects of Optimally Distributed Public Programs. American Economic Review, 76(3): 470-87.

Schultz, T.P. 1988. Educational Investments and Returns. In H. Chenery and T.N. Srinivasan (eds.) Handbook of Development Economics, Volume 1, Amsterdam: North Holland Press.

Schaffner, J. 2004. Determinants of Schooling Investments among Primary School aged Children in Ethiopia. Africa Region Human Development, working paper series.

Singh, R.D. 1992. Underinvestment, Low Economic Returns to Education, and the Schooling of Rural Children: Some Evidence from Brazil. Economic Development and Cultural Change, 40(3): 645-664.

Staiger, D. J.H. Stock. 1997. Instrumental Variables Regression with Weak Instruments. Econometrica, 65(May): 557-86.

Stock, J. and M. Yogo. 2004. Testing for Weak Instruments in Linear IV Regression. mimeo, Harvard University.

Strauss, J. and D. Thomas. 1995. Human Resources: Empirical Modeling of Household and Family Decisions. Handbook of Development Economics, vol. 3, edited by Jere R. Behrman and T.N. Srinivasan.

Strauss, J. and D. Thomas. 2008. Health Over the Life Course, Handbook of Development Economics, vol. 4, edited by T. Paul Schultz and John Strauss eds., Amsterdam: North Holland Press.

Tansel, A. 1997. Schooling Attainment, Parental Education, and Gender in Cote d'Ivoire and Ghana. Economic Development and Cultural Change, vol. 45(4):825-856.

Tansel, A. 1998. Determinants of School Attainment of Boys and Girls in Turkey, working paper, Economic Growth Center, Yale University.

White, H. 1980. A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity. Econometrica, 48: 817-38.

Wooldridge, J. 2002. Econometric Analysis of Cross Section and Panel Data, Cambridge: MIT Press.

Wooldridge, J. M. 2003. Cluster-sample Methods in Applied Econometrics. American Economic Review, 93: 133-138.

Figure 1: Male enrollment rate (%)

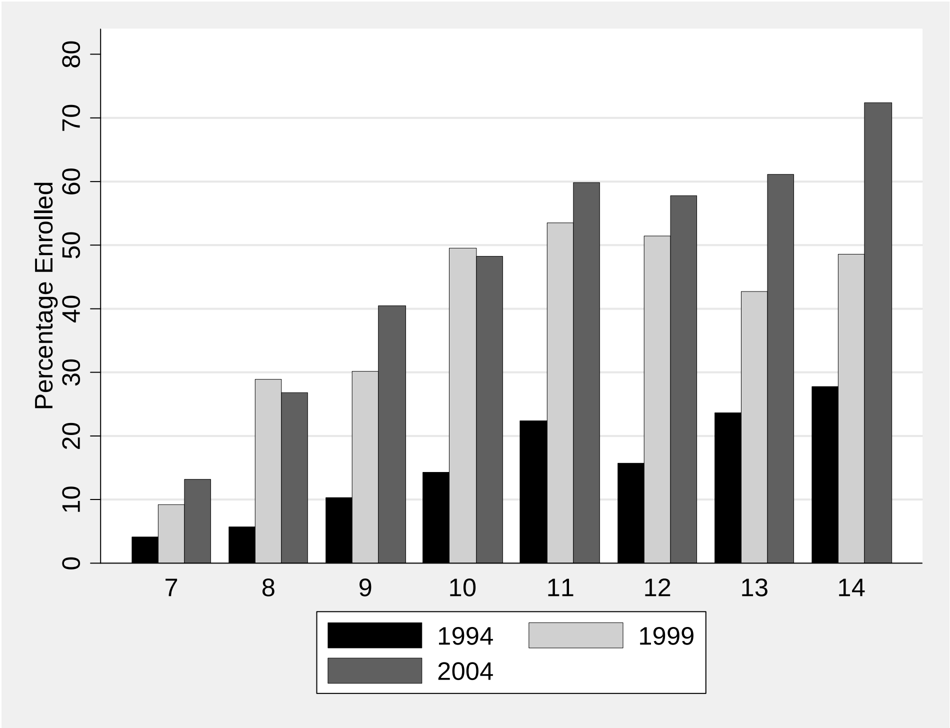

Figure 2: Female enrollment rate (%)

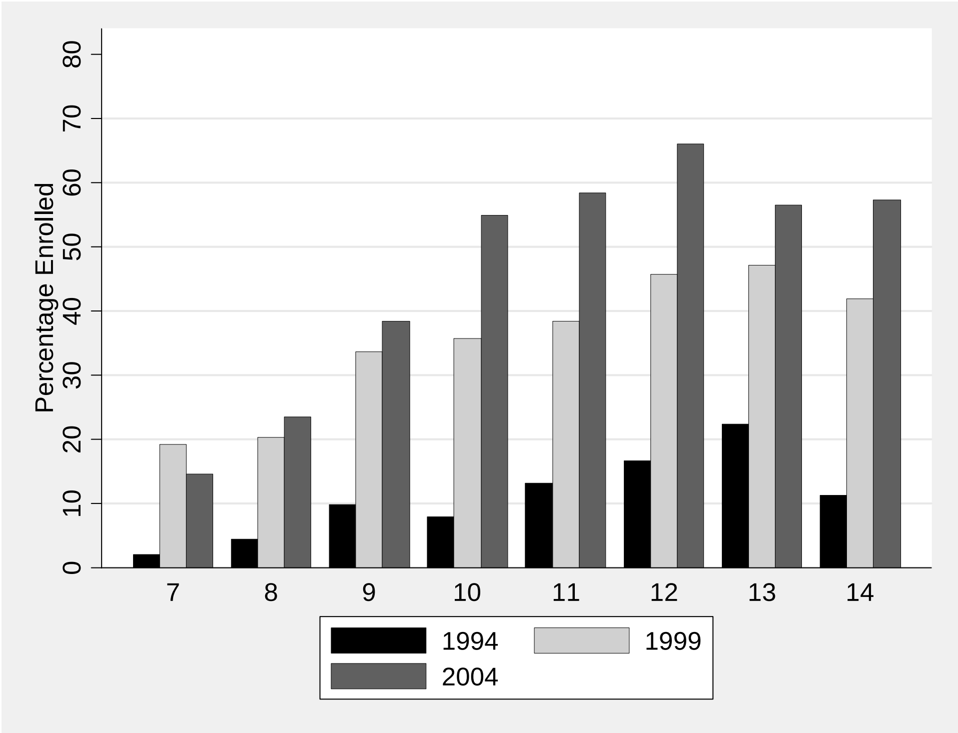

Figure 3: Male relative grade attainment

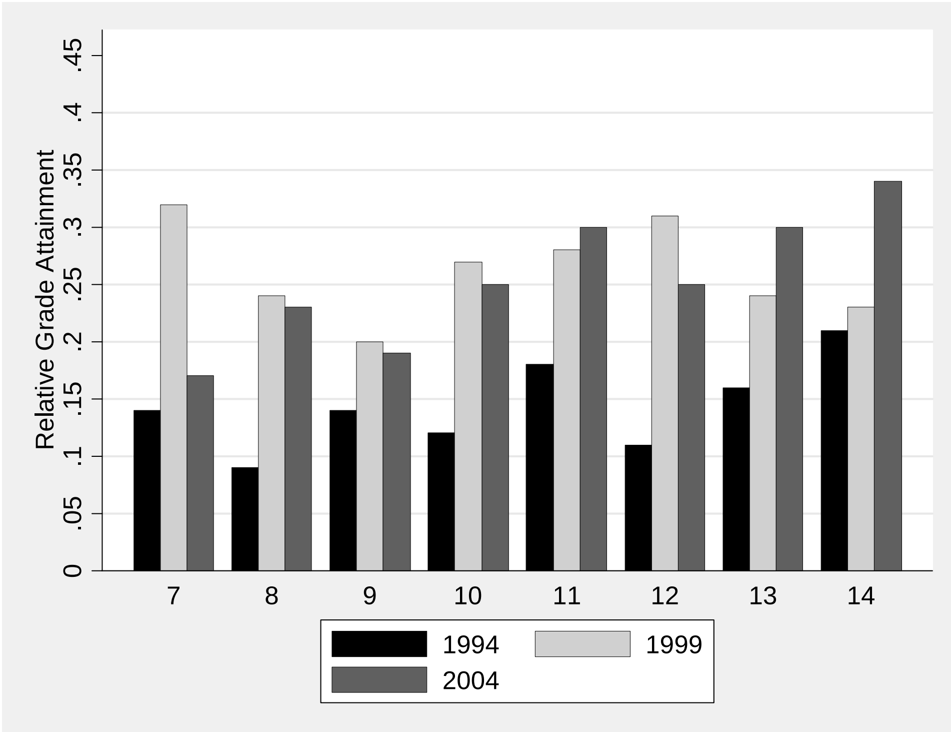

Figure 4: Female relative grade attainment

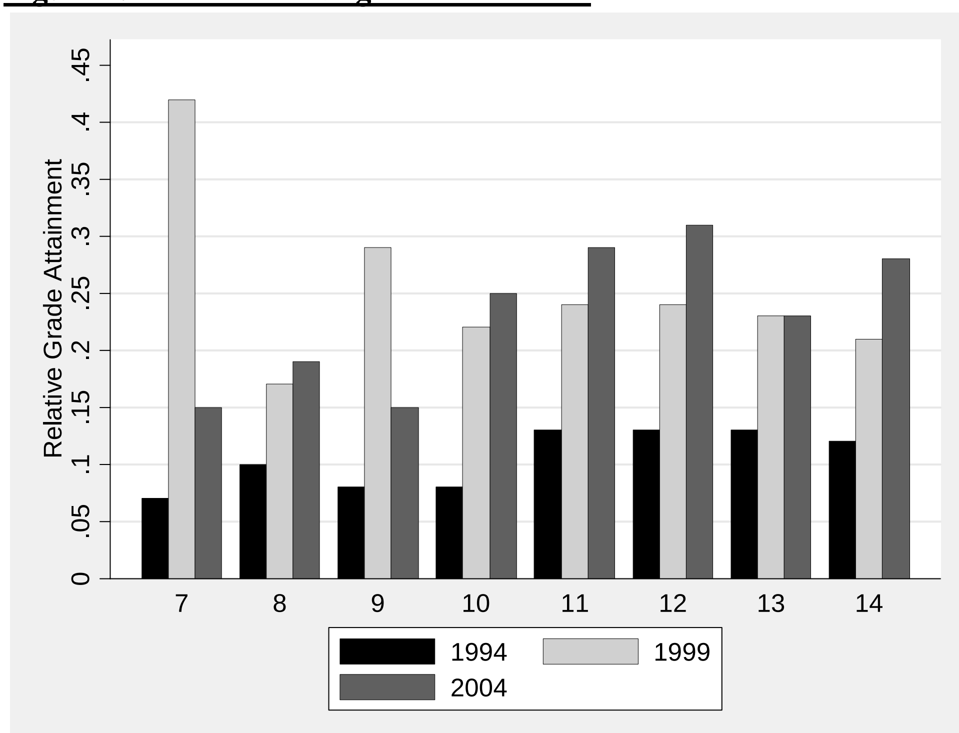

Table 1: Sample Average

Variable	Mean Std. dev	Mean Std. dev	Mean Std. dev
Enroll, Enrolled=1 if currently enrolled in school and 0 otherwise	0.13 (0.34)	0.38 (0.48)	0.45 (0.49)
Completed grades of schooling	0.60 (1.43)	1.13 (1.70)	1.17 (1.67)
Relative grade attained*	0.13 (0.38)	0.27 (0.47)	0.24 (0.41)
Household size	8.29 (3.17)	7.30 (2.90)	7.15 (2.33)
Log real per capita household consumption expenditure (PCE)	3.79 (0.73)	4.05 (0.76)	4.05 (0.75)
Mother's schooling	0.17 (0.37)	0.22 (0.42)	0.31 (0.46)
Father's schooling	0.39 (0.49)	0.44 (0.49)	0.51 (0.50)
Male dummy	0.50 (0.50)	0.50 (0.50)	0.51 (0.50)
Age in years	10.80 (2.29)	10.81 (2.21)	10.70 (2.41)
Land in hectares per adult member	0.57 (0.56)	0.49 (0.50)	0.64 (0.56)
Livestock units	3.16 (4.12)	3.32 (3.10)	3.42 (3.61)
No. of adult male	1.65 (1.18)	1.58 (1.15)	1.49 (0.95)
No. of adult female	1.71 (1.05)	1.70 (1.03)	1.50 (0.83)
Mother's age	38.64 (9.60)	39.76 (9.91)	40.05 (8.65)
Age of the head of the household	48.78 (13.49)	49.27 (12.64)	40.05 (8.65)
Observations	2047	1877	1629

*Relative grade attained = actual grade completed/potential grade.

Table 2: Coefficient estimates on log (PCE) reported in appendix tables A1-A3

Coefficient estimates on log (PCE)	1994	1999	2004
Without IV, column 1	0.03***	0.04**	0.06***
With IV, column 3	-0.01	0.27**	0.17**
With IV, column 4, including actual supply characteristics in the right hand side			0.16***

***-significant at 1%, ** - significant at 5% and *-significant at 10%

Table 3: Coefficient estimates on log (PCE) reported in appendix tables A4-A6

Coefficient estimates on log (PCE)	1994	1999	2004
Without IV, column 1	0.02**	0.04***	0.04**
With IV, column 3	0.10*	0.23*	0.08
With IV, column 4, including actual supply characteristics in the right hand side			0.10**

***-significant at 1%, ** - significant at 5% and *-significant at 10%

Appendix

Table A1: Determinants of schooling enrollment -1994

Covariates	(1) OLS enroll	(2) OLS enroll	(3) IV enroll
Mother's schooling	0.0852**	0.0931**	0.0985**
	(0.03)	(0.03)	(0.04)
Father's schooling	0.1079***	0.1088***	0.1072***
	(0.02)	(0.02)	(0.02)
Log (real pce)	0.0319***		-0.0172
	(0.01)		(0.05)
Land		-0.0218	
		(0.01)	
Livestock units		0.0018	
		(0.002)	
Male dummy	0.0175	0.0147	0.0160
	(0.02)	(0.02)	(0.02)
dummy=1 if the child is 8 years	0.0299	0.0298	0.0300
	(0.02)	(0.02)	(0.02)
dummy=1 if the child is 9 years	0.0833***	0.0792***	0.0799***
	(0.02)	(0.02)	(0.02)
dummy=1 if the child is 10 years	0.0589**	0.0587**	0.0591**
	(0.02)	(0.02)	(0.02)
dummy=1 if the child is 11 years	0.0936***	0.0900***	0.0909***
	(0.03)	(0.03)	(0.03)
dummy=1 if the child is 12 years	0.1374***	0.1380***	0.1398***
	(0.03)	(0.03)	(0.03)
dummy=1 if the child is 13 years	0.2042***	0.1995***	0.2014***
	(0.03)	(0.03)	(0.03)
dummy=1 if the child is 14 years	0.0922***	0.0920***	0.0910***
	(0.03)	(0.03)	(0.03)
Number of adult males	0.0141	0.0130	0.0159*
	(0.008)	(0.009)	(0.008)
Number of adult females	-0.0175**	-0.0207**	-0.0201**

	(0.008)	(0.008)	(0.008)
Mother's age	0.0010	0.0010	0.0011
	(0.0008)	(0.0008)	(0.0008)
Age of the head of the household	-0.0001	-0.0001	-0.0002
	(0.0005)	(0.0006)	(0.0006)
Observations	2047	2047	2047
Village fixed effects	Yes	Yes	Yes
F statistic on the excluded instruments from the first stage regressions			11.17 (0.00)
Hansen J statistic			4.20 (0.24)

Notes:

- Robust standard errors in parentheses
- *** significant at 1%; ** significant at 5%; * significant at 10%
- p-values are reported for the F statistic and the Hansen J statistic
- In column 3, pce is instrumented with land, livestock units, rainfall*land
- omitted age dummy – 7 years, omitted sex dummy – female
- age interacted sex dummies are suppressed and available upon request

Table A2: Determinants of schooling enrollment - 1999

Covariates	(1) OLS	(2) OLS	(3) IV
	enroll	enroll	enroll
Mother's schooling	0.0620	0.0613	0.0413
	(0.03)	(0.03)	(0.04)
Father's schooling	0.0930***	0.0967***	0.0717**
	(0.02)	(0.02)	(0.03)
Log (real pce)	0.0417**		0.2712**
	(0.01)		(0.13)
Land		-0.0383	
		(0.02)	
Livestock units		0.0124***	
		(0.004)	
Male dummy	-0.0689	0.0594	-0.0748
	(0.04)	(0.05)	(0.05)
dummy=1 if the child is 8 years	0.0568	0.0298	0.06061
	(0.05)	(0.02)	(0.05)
dummy=1 if the child is 9 years	0.1824***	0.1892***	0.1892***
	(0.05)	(0.05)	(0.05)
dummy=1 if the child is 10 years	0.2021***	0.2103***	0.1749***
	(0.05)	(0.05)	(0.05)
dummy=1 if the child is 11 years	0.2583***	0.2684***	0.2187***
	(0.05)	(0.05)	(0.06)
dummy=1 if the child is 12 years	0.3124***	0.3126***	0.3118***
	(0.05)	(0.05)	(0.05)
dummy=1 if the child is 13 years	0.3380***	0.3434***	0.3131***
	(0.05)	(0.05)	(0.05)
dummy=1 if the child is 14 years	0.3008***	0.3074***	0.2805***
	(0.05)	(0.05)	(0.05)
Number of adult males	-0.0032	-0.0173 (0.01)	0.0138
	(0.01)		(0.01)
Number of adult females	0.0087	-0.0019 (0.01)	0.0255*
	(0.01)		(0.01)
Mother's age	-0.0035***	-0.0035***	-0.0040***
	(0.001)	(0.0001)	(0.001)
Age of the head of the household	0.0001	0.0002	-0.0003
	(0.001)	(0.001)	(0.001)

Observations	1877	1877	1877
Village fixed effects	Yes	Yes	Yes
F statistic on the excluded instruments from the first stage regressions			9.95 (0.00)
Hansen J statistic			8.97 (0.01)

Notes:

- Robust standard errors in parentheses
- *** significant at 1%; ** significant at 5%; * significant at 10%
- p-values are reported for the F statistic and the Hansen J statistic
- In column 3, pce is instrumented with land, livestock units, rainfall*land
- omitted age dummy – 7 years, omitted sex dummy – female
- age interacted sex dummies are suppressed and available upon request

<u>**Table A3: Determinants of schooling enrollment – 2004**</u>

Covariates	**(1) OLS enroll**	**(2) OLS enroll**	**(3) IV enroll**	**(4) IV enroll**
Mother's schooling	0.0775** (0.03)	0.0839** (0.03)	0.0646* (0.03)	0.1014*** (0.03)
Father's schooling	0.0775** (0.03)	0.0800*** (0.03)	0.0649** (0.03)	0.0843*** (0.02)
Log(real pce)	0.0608*** (0.01)		0.1780** (0.08)	0.1641*** (0.05)
Land		0.0477* (0.02)		
livestock units		0.0055 (0.004)		
Male dummy	0.0099 (0.04)	0.0030 (0.04)	0.0238 (0.04)	0.0196 (0.04)
dummy=1 if the child is 8 years	0.1207** (0.05)	0.1225** (0.05)	0.1180** (0.05)	0.1172** (0.05)
dummy=1 if the child is 9 years	0.2474*** (0.05)	0.2471*** (0.05)	0.2558*** (0.05)	0.2611*** (0.05)
dummy=1 if the child is 10 years	0.4152*** (0.05)	0.4107*** (0.05)	0.4201*** (0.05)	0.4223*** (0.05)
dummy=1 if the child is 11 years	0.4318*** (0.06)	0.4274*** (0.06)	0.4443*** (0.06)	0.4488*** (0.06)
dummy=1 if the child is 12 years	0.5288*** (0.05)	0.5213*** (0.05)	0.5442*** (0.05)	0.5540*** (0.05)
dummy=1 if the child is 13 years	0.4588*** (0.06)	0.4510*** (0.06)	0.4650*** (0.06)	0.4569*** (0.06)
dummy=1 if the child is 14 years	0.4403*** (0.05)	0.4389*** (0.05)	0.4469*** (0.05)	0.4553*** (0.05)
Number of adult males	0.0010 (0.01)	0.0003 (0.01)	0.0005 (0.01)	0.0004 (0.01)
Number of adult females	0.0298* (0.01)	0.0289* (0.01)	0.0420** (0.01)	0.0600*** (0.01)
Mother's age	-0.0022 (0.001)	-0.0019 (0.001)	-0.0022 (0.001)	-0.0018 (0.001)
Age of the head of the household	-0.0003 (0.001)	-0.0002 (0.001)	-0.0010 (0.001)	-0.0006 (0.001)
Distance to primary school in km				-0.0204 (0.01)

Dummy=1 if the village has electricity				0.0409 (0.05)
Dummy=1 if the village as piped water				0.0027 (0.02)
Observations	1629	1629	1629	1629
Village fixed effects	Yes	Yes	Yes	No
F statistic on the excluded instruments from the first stage regressions			32.03 (0.00)	72.28 (0.00)
Hansen J statistic			0.82 (0.66)	0.25 (0.88)

Notes:

- Robust standard errors in parentheses
- *** significant at 1%; ** significant at 5%; * significant at 10%
- p-values are reported for the F statistic and the Hansen J statistic
- In columns 3 and 4, pce is instrumented with land, livestock units, rainfall*land
- omitted age dummy – 7 years, omitted sex dummy – female
- age interacted sex dummies are suppressed and available upon request

Table A4: Determinants of relative grades attained (RGA) – 1994

Covariates	**(1) OLS**	**(2) OLS**	**(3) IV**
	RGA	**RGA**	**RGA**
Mother's schooling	0.0738	0.0820*	0.05204
	(0.04)	(0.04)	(0.04)
Father's schooling	0.1135***	0.1162***	0.1147***
	(0.02)	(0.02)	(0.03)
Log (real pce)	0.0242**		0.1046**
	(0.01)		(0.04)
Land		0.0032	
		(0.01)	
Livestock units		0.0050**	
		(0.02)	
Male dummy	0.0565	0.0552	0.0613
	(0.07)	(0.07)	(0.07)
dummy=1 if the child is 8 years	0.0356	0.0368	0.0368
	(0.04)	(0.04)	(0.04)
dummy=1 if the child is 9 years	0.0091	0.0090	0.0145
	(0.04)	(0.04)	(0.04)
dummy=1 if the child is 10 years	0.0046	0.0068	0.0042
	(0.03)	(0.03)	(0.03)
dummy=1 if the child is 11 years	0.0198	0.0211	0.0242
	(0.04)	(0.04)	(0.04)
dummy=1 if the child is 12 years	0.0426	0.0421	0.0387
	(0.04)	(0.04)	(0.04)
dummy=1 if the child is 13 years	0.0566	0.0524	0.0612
	(0.04)	(0.04)	(0.04)
dummy=1 if the child is 14 years	0.0404	0.0426	0.0423
	(0.04)	(0.04)	(0.04)
Number of adult males	0.0022	0.00007	-0.0007
	(0.007)	(0.007)	(0.007)
Number of adult females	-0.0173*	-0.0191**	-0.0132
	(0.009)	(0.009)	(0.009)
Mother's age	0.0010	0.0011	0.0007
	(0.0009)	(0.0009)	(0.0009)

Age of the head of the household	0.0001	0.0002	0.0003
	(0.0005)	(0.0005)	(0.0005)
Observations	2047	2047	2047
Village fixed effects	Yes	Yes	Yes
F statistic on the excluded instruments from the first stage regressions			11.17 (0.00)
Hansen J statistic			5.34 (0.14)

Notes:

- Robust standard errors in parentheses
- *** significant at 1%; ** significant at 5%; * significant at 10%
- p-values are reported for the F statistic and the Hansen J statistic
- In column 3, pce is instrumented with land, livestock units, rainfall*land
- omitted age dummy – 7 years, omitted sex dummy – female
- age interacted sex dummies are suppressed and available upon request

Table A5: Determinants of relative grades attained (RGA) – 1999

Covariates	**(1) OLS**	**(2) OLS**	**(3) IV**
	RGA	**RGA**	**RGA**
Mother's schooling	0.0061	0.0062	-0.0108
	(0.03)	(0.03)	(0.04)
Father's schooling	0.1066***	0.1106***	0.0892***
	(0.03)	(0.03)	(0.03)
Log (real pce)	0.0452***		0.2322*
	(0.01)		(0.12)
Land		-0.0301*	
		(0.01)	
Livestock units		0.0115***	
		(0.004)	
Male dummy	-0.0916 (0.12)	-0.0888	-0.0963
		(0.12)	(0.12)
dummy=1 if the child is 8 years	-0.2256***	-0.2231***	-0.2225***
	(0.08)	(0.08)	(0.08)
dummy=1 if the child is 9 years	-0.1166	-0.1100	-0.1284
	(0.08)	(0.08)	(0.08)
dummy=1 if the child is 10 years	-0.1999**	-0.1916**	-0.2220***
	(0.08)	(0.08)	(0.08)
dummy=1 if the child is 11 years	-0.1603**	-0.1498*	-0.1925**
	(0.08)	(0.08)	(0.08)
dummy=1 if the child is 12 years	-0.1806**	-0.1802**	-0.1812**
	(0.08)	(0.08)	(0.08)
dummy=1 if the child is 13 years	-0.1680**	-0.1623**	-0.1883**
	(0.07)	(0.07)	(0.07)
dummy=1 if the child is 14 years	-0.1948**	-0.1881**	-0.2113***
	(0.08)	(0.08)	(0.08)
Number of adult males	0.0176	0.0045	0.0314**
	(0.01)	(0.01)	(0.01)
Number of adult females	-0.0251**	-0.0350***	-0.0114
	(0.01)	(0.01)	(0.01)
Mother's age	-0.0002	-0.0002	-0.0006

	(0.007)	(0.006)	(0.01)
Age of the head of the household	0.0019	0.0021	0.0016
	(0.001)	(0.001)	(0.001)
Observations	1877	1877	1877
Village fixed effects	Yes	Yes	Yes
F statistic on the excluded instruments from the first stage regressions			9.95 (0.00)
Hansen J statistic			5.80 (0.05)

Notes:

- Robust standard errors in parentheses
- *** significant at 1%; ** significant at 5%; * significant at 10%
- p-values are reported for the F statistic and the Hansen J statistic
- In column 3, pce is instrumented with land, livestock units, rainfall*land
- omitted age dummy – 7 years, omitted sex dummy – female
- age interacted sex dummies are suppressed and available upon request

Table A6: Determinants of relative grades attained (RGA) – 2004

Covariates	(1) OLS RGA	(2) OLS RGA	(3) IV RGA	(4) IV RGA
Mother's schooling	0.0812**	0.0857**	0.0763*	0.0916**
	(0.03)	(0.03)	(0.03)	(0.03)
Father's schooling	0.0467	0.0491	0.0418	0.0455
	(0.03)	(0.03)	(0.03)	(0.03)
Log(real pce)	0.0417**		0.0869	0.1098**
	(0.01)		(0.07)	(0.04)
land		0.0248		
		(0.02)		
livestock units		0.0028		
		(0.003)		
Male dummy	0.0281	0.0315	0.0366	0.0453
	(0.07)	(0.07)	(0.07)	(0.07)
dummy=1 if the child is 8 years	0.0547	0.0559	0.0537	0.0645
	(0.07)	(0.07)	(0.07)	(0.07)
dummy=1 if the child is 9 years	0.0012	0.0003	0.0045	0.0107
	(0.06)	(0.06)	(0.06)	(0.06)
dummy=1 if the child is 10 years	0.1057	0.1029	0.1076*	0.1138*
	(0.06)	(0.06)	(0.06)	(0.06)
dummy=1 if the child is 11 years	0.1141	0.1107	0.1189	0.1314*
	(0.07)	(0.07)	(0.07)	(0.07)
dummy=1 if the child is 12 years	0.1599**	0.1546**	0.1658**	0.1764***
	(0.06)	(0.06)	(0.06)	(0.06)
dummy=1 if the child is 13 years	0.0892	0.0846	0.0916	0.0989
	(0.06)	(0.06)	(0.06)	(0.06)
dummy=1 if the child is 14 years	0.1291**	0.1278**	0.1316**	0.1355**
	(0.06)	(0.06)	(0.06)	(0.06)
Number of adult males\	-0.0061	-0.0064	-0.0064	-0.00876
	(0.01)	(0.01)	(0.01)	(0.01)
Number of adult females	0.0082	0.0067	0.0129	0.0225*
	(0.01)	(0.01)	(0.01)	(0.01)
Mother's age	-0.0013	-0.0011	-0.0010	-0.0010
	(0.002)	(0.002)	(0.002)	(0.002)
Age of the head of the household	0.0011	0.0007	0.0008	0.0008
	(0.001)	(0.0009)	(0.0009)	(0.0009)

Distance to primary school in km				-0.0214* (0.01)
Dummy=1 if the village has electricity				0.1273** (0.06)
Dummy=1 if the village as piped water				-0.0093 (0.02)
Observations	1629	1629	1629	1629
Village fixed effects	Yes	Yes	Yes	No
F statistic on the excluded instruments from the first stage regressions			32.03 (0.00)	72.28 (0.00)
Hansen J statistic			0.82 (0.66)	0.25 (0.88)

Notes:

- Robust standard errors in parentheses
- *** significant at 10%; ** significant at 5%; * significant at 1%
- p-values are reported for the F statistic on the excluded instruments and the Hansen J statistic
- In columns 3 and 4, pce is instrumented with land, livestock units, rainfall*land
- omitted age dummy – 7 years, omitted sex dummy – female
- age interacted sex dummies are suppressed

<u>Table A7: First-stage regression estimates for the preferred estimates reported in tables A1-A6</u>

Covariates	**(1) IV Enroll RGA 1994**	**(2) IV Enroll RGA 1999**	**(3) IV Enroll RGA 2004**	**(4) IV Enroll RGA 2004**
Mother's schooling	0.2838***	0.0837*	0.1167**	0.2233***
	(0.06)	(0.04)	(0.05)	(0.05)
Father's schooling	0.0026	0.0939***	0.0948*	0.0881*
	(0.03)	(0.03)	(0.04)	(0.04)
Land	0.0664	0.2459	-0.0948	0.1166*
	(0.10)	(0.17)	(0.08)	(0.06)
Land*rainfall	0.0002	-0.0008	0.0009**	-0.0001
	(0.0006)	(0.001)	(0.0004)	(0.0003)
livestock units	0.0287***	0.0272***	0.0466***	0.0612***
	(0.005)	(0.005)	(0.005)	(0.005)
Male dummy	-0.0422	0.0461	-0.1077	-0.1426
	(0.07)	(0.07)	(0.08)	(0.08)
dummy=1 if the child is 8 years	0.0188	0.0012	0.0115	0.0245
	(0.07)	(0.07)	(0.08)	(0.09)
dummy=1 if the child is 9 years	-0.0436	0.0684	-0.0500	-0.0506
	(0.07)	(0.07)	(0.09)	(0.09)
dummy=1 if the child is 10 years	0.0269	0.1243*	-0.0381	-0.0349
	(0.07)	(0.07)	(0.08)	(0.08)
dummy=1 if the child is 11 years	-0.0155	0.1814**	-0.1084	-0.0543
	(0.08)	(0.07)	(0.09)	(0.10)
dummy=1 if the child is 12 years	0.0536	0.0106	-0.1143	-0.0937
	(0.07)	(0.07)	(0.08)	(0.09)
dummy=1 if the child is 13 years	-0.0510	0.1136	-0.0502	-0.0422
	(0.08)	(0.07)	(0.08)	(0.08)
dummy=1 if the child is 14 years	-0.0021	0.0975	-0.0400	-0.0227
	(0.08)	(0.08)	(0.08)	(0.09)
Number of adult males	0.0185	-0.0817***	-0.0368*	-0.0286
	(0.01)	(0.01)	(0.01)	(0.01)
Number of adult females	-0.0524***	-0.0732***	-0.1059***	-0.0789***
	(0.01)	(0.01)	(0.02)	(0.02)
Mother's age	0.0036**	0.0022	0.0014	0.0039*
	(0.001)	(0.001)	(0.002)	(0.002)

Age of the head of the household	-0.0014 (0.001)	0.0015 (0.001)	0.0048*** (0.001)	0.0036** (0.001)
Distance to primary school in km				0.0908*** (0.01)
Dummy=1 if the village has electricity				0.6168*** (0.05)
Dummy=1 if the village as piped water				0.0889** (0.04)

Notes:

- Robust standard errors in parentheses
- *** significant at 10%; ** significant at 5%; * significant at 1%
- p-values are reported for the F statistic on the excluded instruments and the Hansen J statistic
- In all the columns PCE is instrumented with land, livestock units, and rainfall*land
- omitted age dummy – 7 years, omitted sex dummy – female
- age interacted sex dummies are suppressed
- column (1) reports the first-stage regressions corresponding to the preferred IV estimates reported in column 3 of tables A1 and A4.
- column (2) reports the first-stage regressions corresponding to the preferred IV estimates reported in column 3 of tables A2 and A5.
- column (3) reports the first-stage regressions corresponding to the preferred IV estimates reported in column 3 of tables A3 and A6.
- column (4) reports the first-stage regressions corresponding to the preferred IV estimates reported in column 4 of tables A3 and A6.