THE DETERMINANTS OF PRIMARY SCHOOL ENROLLMENT IN ETHIOPIA: EVIDENCE FROM THREE HOUSEHOLD SURVEYS[1]

Julie Schaffner[2]

Abstract

Ethiopia has made tremendous progress in expanding primary education since the mid-1990s. Yet the country continues to face daunting challenges, as it works to meet the goal of primary education for all by 2015. Given the limited resources available for meeting this challenge, it is imperative that resources be spent knowledgably and wisely. This paper aims to inform policy design by examining the lessons to be learned from three recent Ethiopian household surveys regarding the key barriers to primary schooling, and discussing the implications of the results for policy. All three datasets employ large, nationally representative samples, but bring differing strengths to education policy analysis. The Welfare Monitoring Survey/Household Income and Consumption Expenditure Survey of 1999/2000 contains a rich set of variables describing households' distance from various types of economic and social infrastructure, allowing assessment of the role that geographic barriers play in primary school enrollment. It is also the only dataset to contain a good measure of household consumption expenditure. Thus the roles that low income and geographic distance from school play in preventing children from attending primary school may be examined simultaneously. The Labor Force Survey of 1999 allows of children's involvement in work as well as school, shedding some light on the opportunity costs of children's time. The Demographic and Health Survey of 2000 allows more detailed examination of the potential importance of language and parental attitudes in explaining differences in enrollment rates. The research generates both policy-relevant insights, and suggestions for modifications to future Ethiopian data collection efforts that would enhance the ability to draw inferences of relevance to education policy.

---

[1] * The author gratefully acknowledges help and guidance from Andrew Dabalen, Stefan Dercon, Anjini Kochar, and especially Jee-Peng Tan, who commissioned the analysis presented in this paper as an input to a larger study on education in Ethiopia (World Bank 2004). Financial support from the Norwegian Education Trust Fund managed by the Africa Region of the World Bank is gratefully acknowledged.

[2] The Fletcher School, Tufts University

Ethiopia has made tremendous progress in expanding primary education since the mid-1990s. Yet the country continues to face daunting challenges, as it works to meet the goal of primary education for all by 2015. Given the limited resources available for meeting this challenge, it is imperative that resources be spent knowledgably and wisely. This paper aims to inform policy design by examining the lessons to be learned from three recent Ethiopian household surveys regarding the key barriers to primary schooling, and discussing the implications of the results for policy. All three datasets employ large, nationally representative samples, but bring differing strengths to education policy analysis.

The goal in the analysis of each dataset is to understand better the "determinants" of whether or not children of primary school age are enrolled in school. A "determinant" is a characteristic or circumstance that varies across children in the sample, causing variation in their parents' assessment of the costs or benefits of schooling, thus driving variation in school enrollment. "Determinants" is a broad term, which encompasses both "supply side" factors, such as the availability and quality of schools, and "demand side" factors, such as household income and parental awareness and appreciation of the benefits of education. Measuring the strength of impact of these potential determinants on school enrollment rates aids analysis of such policy efforts as building schools, offering scholarships, or developing promotional information campaigns.

The next section places the study into context, by describing the level of primary school enrollment rates, and disparities in these rates across major groups, in Ethiopia and a comparison group of Sub-Saharan African countries. Subsequent sections then describe the datasets employed, describe the general approach taken for specifying econometric models, discuss the application of this framework to the three datasets, present the econometric results, discuss their implications for policy, and offer suggestions for regarding modification to Ethiopian household survey data collection efforts that would improve their usefulness for education policy analysis.

Primary School Enrollment in Ethiopia

Table 1, drawn from World Bank (2004), describes in broad brush the challenges Ethiopia faces in extending complete primary education to all children. Though primary school extends through grade 8 in Ethiopia, the table considers only enrollment and completion rates relevant to grade 1 through 6, for comparability to data from other African countries. Even though primary school enrollments rose 280 percent between 1993-4 to 2002-03 (World Bank, 2004), gross enrollment rates remain low, even by Sub-Saharan African standards. They also exhibit some

profound differences across socioeconomic groups. Ethiopian girls lag behind boys in school enrollment slightly more than is typical in the sub-Saharan Africa comparison group. Children in the poorest quintile of the consumption expenditure distribution lag behind those in richer quintiles, though the gap between poorest and richest quintile is somewhat smaller in Ethiopia than in the rest of sub-Saharan Africa. The most profound Ethiopian socioeconomic disparity identified in the table is the extent to which rural enrollment rates fall far short of urban rates in Ethiopia, a gap that is far greater than in the African comparison group.

The gross enrollment and completion rates, such as those reported in Table 1, are measures commonly employed for assessing the reach of school systems. They compare the total numbers of children enrolled in grades 1-6 (or entering grade 1, or completing grade six) -- regardless of their age -- to the numbers of children of the age appropriate for the activity according to the official plan. In Ethiopia, the official age at which a child may start primary school is age 7. Thus gross enrollments in grades 1 through 6 are compared to the number of children aged 7 to 12 years. For proper understanding of these numbers, then, it is important to note that in Ethiopia many children begin school when older than 7. In fact, as we will see, the mean age of children in first grade is over 8.5 years in urban areas and close to 12 years in rural areas. It is not uncommon for children in first grade to be 14 years old. The gross rate of entry to first grade thus tends to overstate the share of children who ever enter first grade. The high age at entry also probably contributes to the observation that 6th grade completion rates are much lower than first grade entry rates.

The Datasets

The first panel of Table 2 describes the size and structure of the three datasets employed below to study the determinants of whether or not children are enrolled in primary school. All three datasets employ standard clustered samples, derived from a two-stage sampling procedure. The first stage of sampling selected a random sample of small geographic units called enumeration areas (EAs), or neighborhoods of around 200 (100) households in urban (rural) areas. In the second stage, random samples of 12 to 35 households were selected from within each EA, as described in the table. I will make use of the clustered structure below, in our efforts to control for community-level determinants of primary school enrollment. It should be noted that the sample frame for all of these datasets excludes the non-sedentary populations concentrated in the regions of Affar and Somali. For details on sample design and data collection, see CSA (2001), CSA (1999) and the DHS documentation at www.measuredhs.com.

The approach taken by each survey to measuring primary school enrollment is described in the second panel of Table 2, while the third panel presents some descriptive statistics. Though the descriptive statistics are not identical across surveys, they are broadly consistent. All confirm the tremendous gap between urban and rural enrollment rates, and the somewhat smaller but still significant gap between enrollments of boys and girls. They also document the high age at which children start school, especially in rural areas.

Conceptual Framework

The formulation and interpretation of econometric models must be guided by a comprehensive conceptual framework, if we are to: avoid as many potential biases as possible, interpret our coefficients judiciously, and understand the relative merits of the three datasets. We must thus articulate carefully a comprehensive model of the forces that cause primary school enrollment choices to vary within a cross section. I describe here the model of primary school enrollment decisions, which I will apply to each of the three datasets below as best I can. It begins with the assumption that households behave as if they decide whether or not to send child members to school in a given year through a rational decision-making process, in which they compare the benefits and costs of sending the child to an additional year of school.

The Benefits of Schooling. The primary potential benefits of schooling relate to improvements in the child's future socioeconomic status, in which the parents may share either directly (as children eventually contribute to household income and support parents in old age) or vicariously. Households (perhaps especially wealthier households) may also place intrinsic value on education for their children, and may derive tangible short-term enrollment benefits in the form of meals provided to children at school or scholarship funds.

The Costs of Schooling. Direct costs of schooling include tuition and fees, the cost of books, uniforms and supplies required by the school, and the cost of transportation, lodging and board required for attending school. The full cost of schooling may be much greater, however. School attendance carries with it the opportunity cost of the child's time, if school attendance reduces the time the child devotes to activities that expand household income. These include not only working for a wage, but working on a family farm or in a family enterprise, and performing child care or housework that frees up other household members to engage in income-generating activities. Sending a child to school may also carry with it risks and psychic costs, if sending children to school exposes them to risk of abduction, other

crimes, or health hazards, or if it would require sending a child into a situation that is not considered socially acceptable.

Poverty and Discounting. Even if the perceived benefits and costs of sending their children to an additional year of schooling were the same for poor and non-poor parents alike, the poor might choose less schooling for their children because they must weight current costs more heavily relative to future benefits. The difference arises because the poor lack savings and access to credit. This means that they cannot take on any direct economic costs of schooling (that are not compensated by scholarships or feeding programs) without reducing their current consumption; and, since their current consumption is already at a low level, they feel these costs more acutely.

Implications for the list of potential determinants. This simple framework suggests that primary school enrollment rates may vary across children in a cross section dataset as a result of variation in (a) the characteristics of the communities in which they live, (b) characteristics of the households in which they live, and (c) their own characteristics. In specifying and interpreting econometric models of primary school enrollment, then, we must identify the characteristics of communities, households and children that are most likely to influence parental perceptions and comparisons of the benefits and costs of schooling, seek to include as comprehensive as possible a list of controls for these characteristics, and take into account the determinants for which controls are not available when weighing the strengths and weaknesses of our estimates and interpreting our results.

Community-level determinants. By "community" I mean the small geographic area in which the child's household is located, and the residents of that area. The primary features of the community that shape parental assessments of schooling benefits and costs (even after controlling for the characteristics of households themselves) are: the accessibility and quality of the nearest primary schools, local economic conditions shaping parental perceptions of the benefits of good schooling, local economic conditions shaping child work opportunities, and local culture shaping attitudes and beliefs about the propriety and value of schooling.

Household-level determinants. School enrollment choices will differ across households within communities, not only because they differ in their level of income or poverty, but also because they differ in the actual and perceived benefits and costs of schooling that they would encounter. Perceived benefits may differ as a result of differing belief about the benefits of schooling, shaped in part by differences in exposure to schooling. The costs may differ across households living at different

distances from the school, with differential access to scholarship programs, feeding programs or fee waivers, or facing different social constraints.

Child-level determinants. Even within households, enrollment decisions may differ across children of different gender, status and age. The tendency observed around the world for girls to receive less schooling than boys derives, according to our conceptual framework, from parents perceiving lower benefits or higher costs of schooling girls. The perceived benefits of schooling may be lower for girls than for boys, if girls are not expected to work in adult jobs in which schooling increases earnings. The perceived costs may be higher for girls than for boys, schooling is less socially acceptable for girls, or if travel to school exposes girls to greater risks to personal safety. Parents' willingness to invest resources in children may also differ across other variables describing children's position in the family. Even in monogamous households, parents may tend to give priority to the eldest or the youngest. In polygamous households, priority may be placed on the children of the first wife. Increasingly, children who have lost parents to conflict or AIDS find themselves in households where their status may differ from that of birth children.

The costs and benefits of schooling a child in the current year depend in a potentially complicated way on the child's current age. It is reasonably common to suspect that the benefits of schooling diminish with child age (after the official start age for primary school) – because younger children have a greater developmental capacity to learn. It is also common to suspect that the opportunity cost of the child's time rises with age, as the child becomes more and more capable of earning adult returns to labor. These trends help explain why children eventually leave school.

In Ethiopia and other sub-Saharan African countries where many children start primary school quite late, the role of age in shaping schooling decisions may be even more complicated. Parents may perceive that the costs of schooling fall with age at early ages, before rising later. Costs may be high for the youngest children because traveling substantial distances to school over unimproved paths may be too physically demanding or dangerous, or because the care of younger siblings is considered the responsibility of 7, 8 and 9 year olds. The benefits of schooling may also be perceived to rise at low ages, before falling later, if parents believe younger children are not yet prepared to learn.

A Dynamic View. The benefits and costs of schooling a child in a given year probably depend on the child's past schooling experience, including how many grades the child has already completed, whether the child attended school last year. Parents may perceive the benefits of schooling to diminish at higher grade levels,

once basic skills of literacy and numeracy have been acquired. Once a child has started school, the benefits of additional schooling may diminish, and the costs rise, if the child has had to stop out for a substantial period of time. This suggests that a child's current school enrollment is the outcome of a dynamic process, in which not only present community and household circumstances have played a role, but also circumstances experienced at earlier ages.

General Econometric Framework

Basic structure and interpretation of estimating equations. I focus on children of official primary school age, 7 to 14 years old, and employ as dependent variable an indicator of whether or not the child is currently enrolled in school, E_i. I relate this dichotomous variable to its potential determinants using a standard probit formulation, in which I assume that

$$\Pr(E_i = 1) = \Phi(\beta_0 + \beta_u U_i + R_i \beta_R + EA_i \beta_{EA} + H_i \beta_H + K_i \beta_K) \qquad (1)$$

where U_i is a dichotomous variable distinguishing urban from rural locations, R_i is a vector of dummies distinguishing the 11 administrative regions (8 geographic regions and 3 city-states) within Ethiopia's governmental structure, EA_i is a vector of enumeration area characteristics (or, in some specifications, EA fixed effects), H_i is a vector of household, household head and parental characteristics, and K_i is a vector of child characteristics. $\Phi(.)$ is the standard normal cumulative distribution function.

Reduced form interpretation. As indicated above, current school enrollment decisions are part of a larger dynamic process. Estimation of a fully specified dynamic model, in which current schooling outcomes are conditioned on past participation in school, is complicated, because past school participation is endogenous. I choose to focus instead on reduced form current schooling equations, in which I do not condition on past school experience. It is important to keep in mind that while the determinants explicitly included on the right hand side of our probit regressions measure current values, they are "picking up" (albeit imperfectly) both past and present influences on schooling choices that contributed to the current outcome. Their coefficients are best understood, then, as reflecting the long-run effect of changing lifetime resource levels and other circumstances.[3]

[3] For characteristics such as household consumption expenditure, which may fluctuate over time, we must recognize that current income – even if precisely measured – will be an imperfect measure of the resources that have influenced a child's schooling outcomes to date. If this measurement error is classical, it will tend to bias the estimated effect of income on schooling outcomes toward zero. We return to this point below.

Treatment of community, household and child characteristics. Application of the conceptual framework to a particular dataset involves identifying as comprehensive as possible a set of controls for the community, household and child characteristics that influence schooling choices. Identification of the household and child characteristics is reasonably straightforward, as household survey data contain direct information on such characteristics (though they may not contain information on as many characteristics as we might like).

Identifying ways of controlling for community characteristics is more challenging. I take two approaches to controlling for community characteristics. The first involves the combination of urban/rural and regional indicators with within-sample estimates of EA characteristics. The second involves the inclusion of EA fixed effects.

Urban/Rural and Regional Indicators. One of the most striking features of school enrollment data in Ethiopia is the profound difference in enrollment rates between urban and rural regions. Somewhat less dramatic, but still important, are differences across broad administrative regions, whose borders are shaped in part by cultural and language groupings. Within our conceptual framework, primary school enrollment rates may differ between rural and urban areas, or across regions, even after controlling for some household and EA characteristics, as a result of average differences across locations in any of the relevant community or household characteristics that are not explicitly controlled for in the regression. In each dataset below, I seek to gain insight into urban-rural and regional differences, by including indicator variables for these geographic distinctions[4], observing their gross association with enrollment rates (by including them alone on the right hand side), and then observing how their apparent effects diminish as I add various explicit household and enumeration area controls to the regressions.

Urban and regional differences are unlikely to disappear in any of the datasets, because our household-level and especially enumeration area-level controls are incomplete. Characteristics of households and communities that remain uncontrolled for explicitly, and that vary systematically across broad geographic regions, will continue to generate significant differences across regions. The characteristics likely to differ systematically across regions for which I have the weakest controls in all three datasets are school supply conditions related to school quality, school fees and the availability of school feeding programs.

[4] Estimating equations include indicators for all regions represented in the sample except the most populous region, Oromiya. Hence the coefficients represent differences between the indicated region and Oromiya.

Within-sample estimates of EA characteristics. The clustered structure of all three datasets allows calculation of within-EA descriptive statistics that shed light on local exposure to schooling (e.g. share of household heads who are literate) or local economic conditions (e.g. share of household heads who have formal sector jobs or who are unemployed). These are noisy measures of community characteristics for two reasons: (1) we have only small samples (of 12 to 35 households) rather than censuses within enumeration areas, and (2) the enumeration areas employed in the survey may not overlap perfectly with the boundaries of the "community" within which a household's schooling decisions are shaped. They are also incomplete controls, because there are sure to be many more features of communities that matter for enrollment decisions than we can control for. Nonetheless, they may provide insights into the roles of policy-relevant community factors. If, for example, community differences in exposure to schooling appear to drive differences in the schooling of children, then further thought should be directed to the potential importance of school promotion campaigns. Similarly, if local economic conditions that improve work opportunities for children draw children out of school, then further policy analytic energy should be directed toward discovering ways (e.g. changes in local school schedules) that reduce the opportunity cost of schooling.[5]

EA Fixed Effects. A second approach to controlling for community characteristics is to include enumeration area "fixed effects." In practical terms, this requires introducing into the probit specification separate indicator variables for all but one of the enumeration areas represented in the sample. This set of dummies completely absorbs all differences in average enrollment rates across regions – no matter what their source – and focuses attention on how differences in household and child characteristics within communities affect enrollment rates. Variation of household and child characteristics within communities cannot be correlated with community characteristics. This method thus potentially improves estimation of the impacts of household-level determinants, by more completely "holding community-level determinants constant". It does not, however, shed light on the specific features of

[5] As pointed out by Manski (1993), the observed association of EA average variables could reflect three quite different types of "community effects": endogenous community effects, in which the propensity for any one child in the community to go to school is a function of the schooling choices of other households in the community; exogenous or contextual effects, in which the propensity for any one child in the community to go to school is a function of exogenous characteristics of the group of households in the community; and correlated effects, in which the propensity for any one child in the community to go to school is a function of an unobserved community characteristic common to all children in the community. While the EA variables I include appear most directly to represent exogenous or contextual effects, they may pick up endogenous and correlated community effects as well. Attempting to identify these separate channels of effect is left for future research.

communities that drive differences in enrollment rates. We thus remain interested in both approaches.[6]

Characteristics of households, household heads and parents. Households differ in their resources, economic opportunities, access to school infrastructure, and their exposure to and beliefs about schooling. We have two motivations for including various sets of household-level control variables. The effects of some are of direct policy interest. For example, strong household resource effects point policymakers' attention to experimentation with fee waiver and scholarship programs, strong impacts of geographic access to school point policymakers' attention to school supply concerns. Controlling for a comprehensive set of household characteristics – even including some controls of little direct interest to policymakers -- is also vital for preventing biases in estimating the effects of other factors.

In each dataset below, I construct the best possible sets of controls for the following household characteristics: household resources and needs, geographic access to school, household demographic structure (which may shape the opportunity cost of the time of primary school aged children, whether by creating needs for child care or modifying the productivity of child labor in income-generating activities), and characteristics of the household head (or the child's parent, where available) associated with attitudes toward schooling, and to the economic opportunities the household faces.

Child Characteristics. For reasons discussed above, I examine how enrollment probabilities differ across boy and girls, across children of different ages, across children of different birth orders, and across birth children and foster children.

Estimation within Sub-Samples. The basic model of equation (1) includes only a linear term for each potential determinant identified. Compelling generalizations to this model would allow for interactions among the determinants, acknowledging the possibility that the size of the impact on school enrollment of one determinant depends upon the level of another. For example, increases in household income may have differential impact on the schooling enrollment of boys and girls, or of children who live near or far from a primary school. I allow conveniently, though incompletely, for such interactions, by examining model estimates for various sub-samples. I am especially interested in potential differences in coefficients according

[6] In fact, we might prefer to examine the effects of community characteristics much more explicitly. We would especially like to be able to examine the impact of specific changes in the features of local school supply. We return to this, below, when we offer suggestions for future data collection.

to several distinctions. Differences in coefficients between rural and urban areas, may indicate the importance of tailoring policies to tackle the low enrollment rates in rural areas. Differences in coefficients between boys and girls may shed light on the types of interventions most likely to reduce gender gaps in schooling enrollment. Difference in coefficients between younger and older children may shed light on why rural Ethiopian children tend to start school so late. Differences in coefficient estimates between households that are physically close to primary schools may reveal how the relative importance of various barriers to schooling might change as more schools are built, alleviating critical supply constraints.

Even after disaggregating regressions across sub-samples at this level, we recognize that the impacts of various determinants on enrollment rates probably differ across communities in differing circumstances. In interpreting the results it is thus important to keep in mind that the probit coefficient estimates represent an average across diverse communities of the effects of any particular determinant. Finding small estimates of these average effects will not rule out the possibility that the determinants matter significantly in certain localized contexts.

Reporting of Estimates and Standard Errors. To facilitate interpretation of the probit results, I report estimated probability derivatives rather than the probit coefficient estimates ($\hat{\beta}$) themselves. The probability derivative associated with a particular right hand side variable, X_j is equal to

$$\frac{\partial \Pr(E_i = 1)}{\partial X_j} = \phi(\hat{\beta}_0 + \hat{\beta}_u U_i + R_i \hat{\beta}_R + C_i \hat{\beta}_C + H_i \hat{\beta}_H + K_i \hat{\beta}_K)\ \hat{\beta}_j$$

where $\hat{\beta}_j$ is the coefficient on a particular regressor of interest, and $\phi()$ is the standard normal probability density function. This indicates the percentage point increase in the probability of enrollment associated with a one-unit increase in X_j. For example, a probability derivative of -.03 associated with a regressor measuring distance to school in kilometers indicates that a one kilometer increase in distance to school is associated with a 3 percentage point reduction in the probability of enrollment. Put another way, it indicates that as we increase distance to school by one kilometer, enrollment rates fall by 3 percentage points.[7] Notice that the probability derivative for any one regressor is a function of the values of all

[7] The definition for the probability derivative presented in the text is relevant to continuous right hand side variables. In the tables below, many of the regressors are dichotomous. For these, we calculate the discrete change in probability associated with changing the indicator from zero to one (while holding other regressors at their means).

regressors. I follow customary procedure in calculating the probability derivatives for the "mean child" by plugging the means for the regressors into this expression.

Unless otherwise indicated, the standard errors reported in the tables are calculated using formulas that account for arbitrary heteroscedasticity as well as the clustering of observations within enumeration areas. Asterisks indicate that the underlying coefficient estimates were significantly different from 0 at the two-tailed five percent level.

Application to the WMS/HICES 1999/2000

Table 3 defines the variables I have constructed from the WMS/HICES in each of the main categories of potential determinants, and compares the means of the variables within rural and urban sub-samples. The main strengths of this dataset for education policy analysis lie in its inclusion of two measures at the household level: the distance of the household (in kilometers) from the nearest primary school, and household consumption expenditure (which I measure on a real, per adult equivalent basis). I focus especially on obtaining good estimates of the impacts of these two variables, and on examining the extent to which differences in these two variables seem to explain the large urban and regional differences in primary enrollment rates.

The distance to school measure serves as a crude but valuable control for local school supply conditions. Distances vary a great deal, and many children live at substantial distance from school, especially in rural areas. According to the WMS, 19.6 percent of children nation-wide live less than one kilometer from the nearest primary school. In rural areas, however, while 15.4 percent live this close to the nearest school, 18.3 percent live 5-6 kilometers away, 11.8 percent live 7 to 12 kilometers away, and 3.2 percent live more than 12 kilometers away. Rural enrollment rates decline from 43.6 percent among children less than a kilometer from school to 8.1 percent among children at least 12 kilometers from the nearest school. Though these distances are measured at the household level, they contain information both about the extent to which the community in which the household resides has access to social infrastructure, and about the household's location within the community.

Our estimates of the distance effects will be most useful for policy analysis, if they truly represent the intrinsic or causal effect of physical distance from school on enrollment rates. If the intrinsic distance effect is large, then reducing the typical distance to school by building more, and more geographically dispersed, schools would be expected to increase school enrollment rates substantially. A mere

correlation between distance and school enrollment arising out of the fact that more remote households are poorer, for example, would not have this implication; building more schools without improving the incomes of remote households would do little to reduce the barriers to schooling implicit such a correlation. It is thus important to control as well as possible for other household and community characteristics that are likely to be correlated with distance from the nearest school.

The most obvious characteristics that might affect schooling outcomes (for reasons unrelated to the physical distance to school) and be correlated with distance to school are the remoteness of the household from markets and other social services. Remoteness along these dimensions is likely to reduce both household income and household exposure to the potential benefits of schooling. In an effort to control for these dimensions of remoteness, I include measures of the household's distance to the nearest market, and to the nearest post office. It is useful to point out that the typical distance to school in rural Ethiopia is smaller than the typical distance to market, and far smaller than the distance to the nearest post office, reflecting that some schools have been built in otherwise very remote places. The three distances are not highly correlated, giving us hope of separately identifying their effects.

In addition to including controls for remoteness from markets and other economic and social activity, in some specification I include EA fixed effects, in an effort to control more thoroughly for remoteness (and more general nature) of the local community. As discussed below, I also devote attention to matters of measurement and functional form in estimating the distance effects.

Household consumption expenditure is thought to be a good measure of longer-term household income level. In a seasonal and uncertain environment, household income is likely to fluctuate a great deal across short reference periods, such as weeks or months, but households are likely to "smooth" their consumption relative to their income, by saving (perhaps in the form of food, animals or durable items) when income is high and dis-saving when income is low. The level of consumption expenditure in recent weeks, then, is likely to give a better picture of typical income on an annual basis than would the level of income in recent weeks. It is thus extremely useful that the WMS can be merged with the HICES, allowing inclusion of this variable.

Several steps must be taken in transforming the total consumption expenditure measure into a useful indicator of household resource levels that is comparable across households in the sample. First, it is important to recognize that price levels may differ across locations, and to deflate nominal consumption expenditure

measures by an index of price level variation across geographic locations. I employ a regional price index calculated by the CSA and reported in WMU (2002). Second, it is useful to recognize that the same total household income represents fewer opportunities to spend on child education in a household where a larger number of members means a greater pressure to spend income on basic needs. I recognize this in a standard way, by dividing total household consumption expenditure by the number of "adult equivalents" (East African scale) in the household. I also include household size independently in the regressions, allowing for the possibility that economies of scale generate the ability for more primary schooling on the same per person income level, when the household is larger. Finally, to allow for flexibility in the functional relation between household consumption expenditure and school enrollment, while still rendering the estimates easy to read and interpret, I create and include a set of dummy variables indicating in which quintile of the country-wide per adult equivalent consumption expenditure distribution the household resides.

For policy purposes, we would like to identify the intrinsic or causal effect of household income level on primary enrollments, in order to gain insight into the potential to spur enrollments through the use of income transfer, fee waiver and scholarship programs. Again, we must seek to prevent biases in our estimation by including controls for other factors that increase enrollment rates, and that are likely to be correlated with household income level. In general we might worry about two potential sources of bias: measurement error and correlation with unobserved characteristics that influence both income and schooling outcomes. Employing a measure of consumption expenditure rather than income for a short reference period removes one important source of measurement error present in many studies (Behrman and Knowles, 1997). Still, the measure may contain error as a measure of the longer-run resources that have shaped the child's schooling experience to date, especially for the older children, for whom current resources represent a smaller share of the relevant history of resources. They may also contain measures related to faulty price deflation. If this measurement error is roughly "classical" (randomly too high for some households and too low for others), then we would expect this to bias estimated income effects toward zero.

Unobserved characteristics that influence both income and schooling, and that thus might be a source of bias, may be divided into two groups: community and household-level unobserved characteristics. Higher income households may live in communities where local resources and traditions contribute both to higher household incomes and better community infrastructure. Income may thus be associated with a variety of unobserved community characteristics that influence schooling choices, such as accessibility and quality of schools, level of school fees, and local traditions

that place value on investing in the future. Distance to school and EA average characteristics provide imperfect control for these community characteristics. EA fixed effects specifications, however, might be expected to control more completely for them.

Household income may also be correlated with unobserved household-level characteristics that influence both income and schooling. Households that place greater emphasis on the importance of children working to augment household income may higher household incomes and be less likely to send their children to school. Failure to control for such preference characteristics might be expected to bias income effects downward. On the other hand, households that place high value on working hard for both present and future might tend to have higher incomes and be more likely to send their children to school. Failure to control for such preference characteristics might be expected to bias income effects upward. While I have no direct controls for such traits, I attempt to include as rich a set of variables describing the household as possible. Perhaps the most important proxy for preferences is the indicator of whether the household head is literate, capturing previous exposure of the household to schooling.[8]

Table 4 summarizes the insights that can be gained from this dataset into the source of differences in enrollment rates between rural and urban areas and across regions. It presents estimates of probability derivatives associated with the urban and regional indicator variables in a variety of specifications. The first includes only the rural/urban indicator, and reflects gross rural-urban differences. The second introduces the regional indicators, demonstrating that the gross rural-urban difference is not just an artifact of differences across regions, some of which are largely urban and some

[8] In principle, we could use instrumental variables methods to resolve problems of both measurement error and endogeneity (arising out of correlation with household-level characteristics that influence both income and schooling). This would require the availability of variables to serve as instruments, which help determine a household's income level, but can reasonably be assumed to influence schooling decisions through no channel other than their influence on income. It is difficult to imagine such variables. Behrman and Knowles (1997) employ IV to estimate the effect of income on schooling, but this is possible only because they are attempting to estimate a different notion of income's effect on schooling outcomes. Here we aim to measure the ceteris paribus effect income on schooling, which is of relevance to the analysis of income transfer policy. Behrman and Knowles are attempting to estimate the more general correlation between income child schooling, which includes not only the ceteris paribus household income effect, but also correlations that arise because higher income households contribute to improved local school infrastructure, and because higher income households tend to have more educated parents who place greater priority on sending children to school. Given this aim, they do not want to include parental education and other characteristics that are correlated with household income in the schooling regression. This frees them to use such variables as instruments.

largely rural. The large rural-urban differences are observed within regions. Subsequent specifications introduce community, household and child level controls.

The addition of the lone EA-level variable, the percentage of household heads within the EA that are literate (and thus have personal exposure to school of some sort) reduces the urban-rural enrollment differential by over 10 percentage points. Including our crude control for school supply differences reduces the remaining urban-rural differential by even more. Other household characteristics – even household poverty levels – seem to play at most a very small role in explaining regional differences in enrollment rates. And even after introducing all these controls, the rural-urban differences remain very large, suggesting that there is more to understanding rural-urban differences in enrollment rates than just differences in physical availability of schools and income levels. Differences in both school quality and desirability, on the one hand, and in cultural understanding of the benefits of schooling, on the other, both stand out as possible contributors to the explanation of remaining rural-urban differences.

Table 5 presents the main estimates of the effects of the household, child and community-level characteristics. (Coefficients on urban and regional indicators are suppressed in this table.) The first specification is a standard probit containing all controls as described in Table 3. The other three specifications are motivated by concern about possible econometric problems in the simple specification. The second column replaces the simple distance variables by truncated distance variables, in which distances over 15 kilometers are truncated to 15. This is motivated by the observation that many households report quite large distances to the nearest primary school (and to other services), and that at sufficiently large distances, an additional kilometer must eventually come to have little effect on primary school enrollment rates. Experimentation with quadratic and spline specifications led me to pick this simple formulation involving the truncated variables, which is equivalent to a spline, in which each kilometer of distance up to 15 kilometers has the same impact on the argument to the probit function, but additional kilometers after 15 have no effect.

The third specification in Table 5 replaces the household-level distance measures by the within-EA medians of the distance measures. The motivation here arises out of concern with measurement error in the household-level reports. Within EAs the reported distances vary a great deal, leading to some suspicion of inaccuracy in the reports. If the measurement errors are roughly independent and identically distributed across households within an EA, however, they should roughly “cancel out” when distances are aggregated for calculating medians within EAs.

The final specification in Table 5 replaces the EA, urban and regional controls with EA fixed effects. We are especially interested in this specification, as we have reason to worry that both the household's distance to the nearest primary school and its consumption expenditure level are correlated with a variety of unobserved community characteristics.

Table 6 demonstrates large and robust effects of the household's distance to the nearest primary school on primary enrollment rates. Each kilometer of distance appears to increase enrollment rates by 2 to 3 percentage points. Many children live 5, 10, and sometimes many more, kilometers from the nearest school. Reducing distance from 10 kilometers to zero, then could be expected to increase enrollment rates by 20 to 30 percentage points.

Household income level effects also appear quite robustly estimated, and are even larger in the community fixed effects specification than in the other specifications. Even so, they are somewhat smaller than we might have expected, given the current level of enthusiasm in school policy circles for income transfer programs aimed at getting children into school. Increasing the incomes of households in the lowest income quintile enough to put them in the second quintile would increase their primary enrollment rates on the order of 4 or 5 percentage points. Increasing their incomes enough to put them in the highest income quintile would increase enrollment rates by only 7 to 14 percentage points.

Several other estimated effects in Table 6 merit note. Exposure to schooling, as measured both by having a household head who is literate, and living in an EA in which the share of household heads who are literate is high, both have quite large apparent effects on enrollment rates. This, combined with the large rural-urban differentials that remain even after controlling for distance to school and household income level, point to a potential role for culture, beliefs and attitudes – and perhaps education promotion campaigns -- in shaping primary enrollment decisions.

Table 7 presents simple and fixed effects probit results run separately for rural and urban areas. It demonstrates that the roles of distance to school, income and the EA characteristics captured by the literacy rate are much stronger in rural areas – where enrollment rates are low – than in urban areas, where primary education is indeed beginning to approach universality. Gender differences in schooling rates are also much larger in rural areas.

Finally, Table 8 presents the results of running the simple probits (without EA fixed effects) separately for boys and girls, younger and older children, and children living

near to and farther from primary schools. Distance matters noticeably more for girls than for boys, and for younger children than for older children, raising hope that improved physical access to school might help reduce gender gaps and reduce the age at first enrollment.

Income effects appear slightly higher for boys than for girls, but for neither group are they very large. Income effects appear substantially larger among the youngest children than among the oldest children. This result is subject to a variety of interpretations, raising more questions than it answers. We might suspect that current consumption expenditure is a more accurate measure of the history of household resource levels that have influenced a child's progress through school to this point for younger children than for older children. If this implies a higher degree of classical measurement error in our measure of permanent income for older children than younger, this could explain the smaller apparent effects on the schooling of older children as a result of classical attenuation bias. The larger effects observed among younger children might then be the more accurate. The effects might also appear larger for younger children, because the recent building of schools renders schooling more of a real possibility for them than it was for their older siblings. Either way, the results offer some sign that further investigation of the potential role of income transfers is merited. On the other hand, even for the younger children, it is only at in the highest income quintiles that the effects of income become important. There is still little evidence that modest increases targeted at the poorest households would increase enrollment rates vary greatly.

Income effects also differ little between households that are close to and far from schools. School enrollment rates in rural areas are low even among households living within two kilometers from school. We might have expected that income plays an important role in determining which households take advantage of local school infrastructure, but these regressions do not provide strong evidence of this.

Literacy of the child's household head and of household heads in the child's EA have apparent effects that are consistently strong across all groups represented in the table.

Application to the LFS 1999

The LFS 1999 is a much larger survey than the WMS/HICES 1999/2000, containing more households per enumeration area, as well as more households overall. Being a more focused labor force survey, it contains less information on living standards and education, but contains more information on children's participation in work activities

as well as in school.[9] I thus focus our attention on deriving insights from it about the opportunity costs of schooling for primary school aged children in Ethiopia. Our approach is to study the effect of our community, household and child characteristics not only on school attendance, but also on involvement in work activities that might conflict with schooling.

The LFS elicited information on child work activities in the following manner. Children who were not attending school were asked: "What were you doing during the last 7 days? The possible responses were (1) For household- agricultural, (2) for household- non-agricultural, (3) paid employment – agricultural, (4) paid employment –nonagricultural, (5) paid domestic service, (6) self employment, (7) unpaid domestic service, and (8) didn't work. Children who were attending school were asked: "What were you doing during the last 7 days in addition to attending school?" and were given the same options. It is thus possible to construct indicators of participation in various kinds of work for all children, whether they attended school or not. The tables below focus on participation in income-generating work ("job" for short), which is defined as involving activities 1, 2, 3, 4,5 or 6 in the code schemes just described, and participation in "house work", which is associated with response code 7. The indicators of "job work" and "house work" are thus constructed in a way that renders them mutually exclusive. They are not, however, constructed as mutually exclusive alternatives to school attendance.

Table 9 presents some simple patterns related to participation in schooling and work by rural and urban boys and girls. Large fractions of children, especially in rural areas, report working in jobs or housework, whether they are in school or not. Apparently much of the work in which they are involved is compatible with school attendance. As is to be expected, boys are more likely to report involvement in directly income-generating work, while girls are more often involved in house work. For boys and girls, in rural and urban areas alike, reports of house work are even higher among those participating in school than the others.

Table 10 demonstrates what happens to estimated urban and regional coefficients in primary school enrollment probits as the various sets of regressors available in this dataset are added to the specification. Again, adding regional controls makes little difference to the large estimated urban-rural differential. Adding EA controls – percentage of household heads in the EA who are literate, who have formal sector

[9] The LFS is also the only survey for which we had adequate geographical information for merging with school census data at the woreda level. Unfortunately, for reasons discussed in Schaffner (2003), merging at the woreda level turned out not to produce many useful results. While we continue to work on improving the merge, in this paper we do not introduce the merge.

jobs and who have no job – reduces the remaining rural-urban differential significantly. Unfortunately, given the absence of even crude controls for school supply, it is not clear whether this is because community-level economic and social conditions are themselves important for explaining the differences, or because the EA variables are correlated with differences in schooling infrastructure, whose effect they are picking up. Again, adding household and child controls makes little difference for the estimated rural-urban differentials.

The estimated effects of household and child variables, in specifications controlling for community effects in two ways (either urban, regional and EA averages, or EA fixed effects) are presented in Table 11. Care must be taken in interpreting their coefficients, given that lack of any direct measures of household income level in this dataset. Simple indicators of whether the household head has a formal sector job (which generally pay better), whether the head has no job, and the share of adults in the household with formal jobs, are the closest I get in this dataset to controlling for the household's income level. As these variables are probably correlated with the degree of development and likely existence of social services in the community, it is especially interesting to observe whether their apparent effects persist in the EA fixed effects specifications.

The interpretation of many coefficients must be quite different here compared to the WMS/HICES case, because this dataset lacks a good household income measure. Whereas the estimated effect of household head literacy in the WMS/HICES dataset is an effect that remains even when income level is held constant, in this dataset it captures effects of household head literacy on primary school enrollment through effects on both income and attitudes or beliefs. Similarly, whereas in the WMS/HICES household structure variables pick up effects independent of their effects on household income, and thus reflect to a greater extent the importance of household structure in shaping the opportunity cost of children's time, in the LFS they pick up effects working through both income and opportunity cost channels.

The urban results in Table 11 suggest that demands for child care for younger siblings present a barrier for the schooling of some children, although the average effect observed in the regressions is not tremendously large. An increase in the share of young children in the households by 20 percentage points, for example, would increase primary school enrollment rates on the order of 4 percentage points. The more significant effect of family structure on school attendance in rural areas is the depressing effect on enrollments of having a larger representation of adult males. This somewhat unexpected effect, which is seen in all three datasets, raises the possibility that children's opportunity time in farm and family enterprise work is higher

when there are adult males to supervise them. In communities EA-level calculations show employment to be more formal, primary enrollments tend to be higher, whether because schooling is more valued, because incomes are higher, or because school infrastructure is better. In EAs where unemployment is more prevalent, suggesting weaker labor markets, schooling enrollment is also higher, suggesting some substitution between schooling and work for children. Interestingly, however, the rate of literacy among household heads in the EA continues to play a strong role in shaping enrollment rates (especially in rural areas) even after including rough controls for labor market conditions.

Table 12 takes advantage of having simultaneous information on children's involvement in school and work, presenting side by side the effects of community, household and child level determinants on three dependent variables describing participation in the three child activities. Having reason to suspect that children's time allocation choices differ between boys and girls, and between rural and urban areas, I differentiate the analysis along those two dimensions.

In many cases variables that have large apparent effects on primary school attendance have much smaller or insignificant effects on work activities, suggesting that in many cases schooling and work are reasonably complementary. For example, household head literacy and EA literacy have much stronger associations with school attendance than with involvement in jobs or housework. This suggests that, at least in part, adult literacy influences schooling choices not by relaxing constraints related to opportunity costs, but by improving parental perceptions of the benefits of schooling or possibly defraying direct costs of school enrollment. Similarly, children who have lost father or mother (or possibly both) are less likely to attend primary school compared to others in similar households and communities, but this is not mirrored by significantly greater participation in housework and income-generating activities, as we might expect if their inferior school attendance derived from their being treated more like servants than children.

Some significant effects on school attendance are mirrored, however, by opposite changes in work activity. For example, an increased representation of young children in urban households reduces school attendance for both boys and girls (though more for girls) and increases reports of participation in housework, as we might expect if caring for younger siblings sometimes prevents children from attending school. This shows up for rural girls as well. The reduction in boys' schooling associated with having more adult males in rural households is also mirrored by an increase in participation in income generating work. The overall impression of the estimates is that while work and child care responsibilities may present barriers for some subsets

of children, there is no general incompatibility between schooling and work. Improving parental understanding of the benefits of schooling, and making schools more accessible, then, have real potential to increase enrollment rates, even when potential opportunity cost barriers cannot be tackled directly.

Application to the DHS 2000

The DHS is the smallest dataset, but contains richer information about the child's parents and siblings, including questions about native language and attitudes. Because many of the most interesting variables are available only for children whose mothers are between 15 and 49 years old and responded to a special female questionnaire, I restrict attention to children between 7 and 14 years old, whose mothers are alive, living in the household, and in the 15-49 year age range. Table 13 defines the variables employed in the analysis, and presents their means in rural and urban sub-samples.

As with the LFS, the DHS lacks a direct measure of the household's income or poverty level. Unlike the LFS, the DHS contains a variety of simple indicators of household assets, which some authors consider as a group to offer reasonable control for a household's wealth level (see, e.g. Filmer and Pritchett). I employ them, but with the following reservation. While having such assets as improved floors, access to electricity, indoor plumbing, land, animals or cash crops are certainly associated with wealth level, they are also profoundly associated with the nature of the community or region in which a household lives. As can be seen in Table 14, urban households have electricity, while for the most part rural households do not. Similarly, rural households have land, while most urban households do not. Even within locations treated as officially "urban" and "rural" by the sampling frame, there is likely to be significant variation across communities in the degree of "urbanicity". Estimated asset effects on enrollment probabilities may then pick up not only the effects of household wealth, but also the effects of community characteristics (related to school infrastructure and labor market conditions) that are associated with the degree of urbanicity. Only the EA fixed effects specifications with the estimated asset effects be purged of these biases arising out of correlation between asset indicators and the urban or rural nature of the community.

Table 14 compares apparent urban and regional effects across specifications introducing increasing numbers of controls. As in the other datasets, urban-rural differences are little affected by controlling for region. Introducing the household asset controls reduces the apparent rural-urban gap considerably, but – for reasons described in the previous paragraph – this is no surprise, and cannot be interpreted

as evidence that differences in wealth between regions play a large role in explaining rural-urban differences in primary school enrollments. Again, adding other household-level controls reduces rural-urban and regional differences by very little.

Table 15 presents estimated household, child and (where relevant) community variable effects, separately for rural and urban areas, and employing our two approaches to controlling for variation in community characteristics. Fixed effects estimates, in which the estimated assets effects are even bigger than in the simple probit estimates, indicate that within communities households with more tangible signs of wealth are significantly more likely to send their children to primary school. Unfortunately, it is difficult to gauge the size of the effect. How much of an income transfer would be required to improve a household's economic circumstances as much as economic circumstances differ between households that do and do not have improved floors?

In rural areas, whether the mother's language is Amarigna (rather than Oromigna or the "left out" category of less prevalent languages than Amarigna, Oromigna and Tigrigna) has a strong association with child school enrollment. Father's exposure to schooling, as measured by having completed grade 1, has the same substantial effect on enrollment rates as was seen in household head literacy effects in other datasets. The two controls for the degree of "modernity" of a household's attitudes – whether the mother listens to the radio, and whether she reports that is it justified for a husband to beat his wife if she goes out without telling me – produce mixed results. The children of mothers who listen to the radio are noticeably more likely to attend school. Attitudes about men's treatment of women, however, picks up only small effects. The regression also produces little evidence of strong birth order effects.

Policy Discussion

The preceding econometric study of three household surveys produces the following conclusions regarding the barriers to primary school enrollment in Ethiopia. First, despite tremendous efforts to extend the reach of the primary schooling system over the last decade in Ethiopia, supply constraints continue to play an important role in preventing some children from attending primary school. The estimated effects of a household's distance to school are large in a wide range of econometric treatments. Estimates imply that building schools in communities that were previously 10 kilometers from the nearest school would increase school enrollment rates by 20 to 30 percentage points. The effects might be even larger for girls, tending to reduce gender differences in enrollment rates, and for younger children, tending to reduce the age at which children begin school. Improving the supply of schools in rural areas

relative to the supply in urban areas could also be expected to narrow significantly the difference in enrollment rates between urban and rural areas.

Second, relaxing supply constraints by building more schools is unlikely to be sufficient, on its own, to draw all children into school. Even among households living within 2 kilometers of the nearest primary school, many children do not attend school, especially in rural areas. These observations point to the potential importance of constraints households may face even when schools are near by. They may be "supply side" constraints, related to the lack of space, lack of quality instruction or failure to offer the desired grades at the local school, or "demand side" constraints, such as those considered in the following paragraphs.

Third, while household income levels play some role in determining primary school enrollment probabilities, there is little evidence that a broad program of modest income transfers would eliminate the remaining barriers to school enrollment. Estimated income effects – even after taking as much care as possible with measurement and adequate controls – are of modest size at best. This does not rule out the possibility that in some local contexts income transfers could have a large impact on enrollment rates, though it does cast doubt on the cost-effectiveness of a widespread income transfer policy. It suggests that if income transfers are to be implemented, it will make most sense to implement them on a narrowly targeted and experimental basis.

Fourth, though schooling is consistent with a wide range of child work activities prevalent in Ethiopia, some types of work do seem to compete with school attendance. The association between the presence of younger siblings and reduced enrollment rates, especially for girls and especially in urban areas, points to the potential value of providing daycare for younger siblings. As in the case of income effects, the estimated effects are at most modest in size, suggesting that any efforts to provide day care should be targeted (in this care, to communities where labor markets and cultural norms are such that care for younger siblings is a significant activity among school-aged children) and accompanied by careful evaluation. The tendencies for rural boys in households with more adult males to engage in more frequently in farm work and less frequently in school, and the tendency for enrollment rates to be higher where higher unemployment rates suggest weaker labor markets, raise questions about local school schedules, and whether modifications might in some cases render schooling and work more compatible.

Fifth, a variety of indirect evidence suggests a potentially strong role for parental exposure to, and beliefs about, the benefits of education in shaping their decisions

regarding the education of their children. Even after controlling for household income and distance to the nearest school, the literacy of a child's parent or household head and the rate of literacy among household heads in the child's neighborhood both have very large estimated effects on primary school enrollment rates. After controlling as well as possible for other household and community characteristics, enrollment rates also remain higher among households that live in urban areas and that undertake relatively "modern" practices like listening to the radio. Taken together, these patterns, though sketchy, raise the possibility that primary school promotion and adult literacy campaigns, which offer adults an enhanced appreciation of what education can offer their children, could play an important role in raising enrollment rates.

A final observation derives more from what the results fail to show than what they show. No one potential determinant plays an overwhelming role in explaining primary school enrollment rates. Many of the estimated impacts are modest in size. Rather than being robust, the sizes of key impacts vary across sub-samples. All this is consistent with the hypothesis that the critical barriers to schooling vary from community to community and group to group. Improving access to schools, offering cash transfers, providing care for younger siblings and other interventions may each have the potential to effect large improvements in enrollment rates for some communities or groups, while having little potential to impact other communities and groups. What we observe in the regressions may be the average of such diverse effects. While further research may shed some light on the conditions under which various interventions would be most effective, these results also suggest the potential importance of targeting and of decentralized selection of approaches to increasing enrollments in specific communities and groups.

Suggestions for Future Data Collection Efforts

Econometric study of the three datasets has produced some useful insights about the barriers to primary schooling that must be overcome if all children are to be drawn into primary schooling; but it does not answer all questions. This section discusses modifications to Ethiopian data collection efforts that would enhance the potential for econometric analysis to shed light on policies for increasing primary school enrollment rates. The dataset that produced the most useful results in the present research was the WMS/HICES, because it contained a good measure of household income (household consumption expenditure) and at least a crude measure of school supply (distance to nearest primary school). I thus couch my suggestions as modifications to future waves of the WMS/HICES.

Addition of a Community Questionnaire. Community characteristics figure importantly into our model of primary school enrollment decisions, but can be captured only indirectly and imperfectly on the basis of data reported by sampled households. It is thus useful to consider the addition of a "community questionnaire" administered in each enumeration area included in the sample, as described in Frankenberg (2000).

Perhaps the greatest weakness in our treatment of community characteristics was the inability to characterize fully local school supply conditions. Inclusion of the "distance to nearest primary school" variable in the WMS/HICES was of great value, but even that is far from a perfect indicator of the availability of schools for the household's children. Many rural schools include only grades 1 through 4, and a few only grades 1 through 6, rather than the full 8 years of primary school, presenting a supply constraint for children who started young and completed those grades (World Bank, 2004). Many schools are also over capacity, and may give the impression that spaces are not available to some local children. Beyond this, even where spaces are available, school quality and the availability of school feeding programs may vary, with strong effects on parental assessments of the value of sending their children to school. A final characteristic of local school supply that would be useful for analysis is how long ago the nearest primary school was built. This would be useful for untangling the extent to which children are starting school at older ages because parents prefer not to send younger children to school, and the extent to which they started late because schools were built only very recently.

A community questionnaire could be used to characterize local school supply conditions. Protocols could be established by which interviewers identify the nearest primary school to the enumeration areas. Interviewers could record the distance from the enumeration area to the school, and possibly characterizing the quality and safety of the roads or paths by which children could get from the EA to the school. They could then allow characterization of the school in one of two ways. If feasible, they could obtain the code by which the nearest school is represented in the school census (EMIS) database (allowing school census data to be merged with the WMS). Alternatively, they could administer their own short list of questions about the school. If the second route is taken, candidate questions about the nearest school would include:

- When was the school established?
- What grades of instruction does it offer?
- How many students, teachers and classrooms does it contain?
- What programs are offered related to school meals, care for younger siblings, scholarships or conditional cash transfers, and key community initiatives that might be taken to encourage school enrollment?
- What is the language of instruction?

In addition to characterizing local school supply, a community questionnaire could be used to characterize local economic conditions, local supply of non-formal education opportunities for adults and children. Improved deflation of nominal consumption expenditure data may also be possible if community questionnaires are used to identify local market prices of standardized qualities and quantities of basic consumption items. For guidance on the development of community questionnaires, see Frankenberg (2000).

Standardization of Wereda Codes Across Data Sources. Local economic conditions of relevance to households' schooling decisions may pertain to somewhat larger geographic areas than those encompassed by the enumeration areas. An obvious geographic unit of interest is the wereda, an administrative unit for which a range of statistics are available from other sources.[10] The numeric codes given to specific weredas are not currently uniform across Ethiopian surveys. Even when keys matching numeric codes to transliterated wereda names are available, matching weredas across datasets is very difficult, because of the wide variety of spellings and renditions for wereda names. When datasets are disseminated without wereda names (as in the case of the WMS/HICES data I worked with), even this imperfect matching is impossible. It would thus be useful to create a list of unique numeric identifiers for all weredas, and to encourage the use of these codes in all data collection efforts. [11]

Improvements or Additions to Household-Level Questions. Several improvements to household level questions suggest themselves. (Some of these would be less important if the survey also contained community or school questionnaires.)

- Consider replacing the single question about distance to nearest primary school with questions regarding the distances to the nearest schools offering grades 1-4, 5-6 and 7-8.
- Develop simple questions to elicit parental assessments of whether spaces are available at the nearest school, and how long ago the school was built.

[10] The wereda is a sub-regional administrative unit that spans both rural and urban areas, and the level of government to which greater authority in the provision of public services like schooling is to be devolved in on-going decentralization efforts in Ethiopia. Wereda popoulations range from about 9000 to about 535,000, with mean of about 139,000 (according to the CSA, 2002).

[11] A cautionary note to researchers attempting to employ merged data. Weredas contains both rural and urban areas. When possible it will be useful to merge data across sources not at the level of the entire wereda, but at the level of rural and urban areas within weredas. I managed to merge LFS survey data with school census data aggregated to the wereda level, not distinguishing rural and urban sub-regions. My assessment of the results was that this probably provided a poor characterization of local school supply conditions for many enumeration areas.

- Refine the current question regarding reasons why households do not make use of the nearest primary school. In the WMS of 1999/2000 nearly half of all responses to this question were coded as "other." Improved categorization of responses would allow more insights to be derived from this question.
- Using qualitative research, design a question about parents' beliefs regarding the value of schooling, perhaps distinguishing the perceived value for boys and girls.
- Consider asking the primary language of the household or household head.

Improvements to Individual-Level Questionnaires. Several improvements that would be useful for studying the determinants of primary school enrollment include:[12]

- For each individual, ask explicitly whether he or she has ever attended school, and if so, what was the highest grade completed. (Currently, individuals are asked whether they are literate, and only those who report themselves as literate are asked the highest grade attained in school.) This will allow more accurate identification of children and adults who have ever attended school, as well as more complete identification of highest grade attained.
- Consider replacing this household-level question about non-use of primary school by an individual-level question to be asked of each child who is not in school: "Why is this child not attending school?"
- For children who have attended school, ask the age at which they first attended primary school, and consider designing a question for children who first attended school after age 7 why they did not start school at age 7.
- For all children, whether in school or not, ask simple questions about work and house work activities, such as those employed in the LFS. LFS results give the impression that the indicators of participation in job work and house work were too broad. A few additional questions might help in refining the picture of children's time use. Children who report any kind of work or house work could be asked whether they perform this work for more or less than some threshold number of hours per week, to aid in identifying children engaged in demanding work responsibilities. All children could also be asked more direct questions about whether they are responsible for the care of younger children.

[12] An additional suggestion that is relevant to the study of repeat rates rather than enrollment rates is this: For children who have attended school last year and are attending school this year, ask explicitly whether they have advanced one grade, are repeating the same grade, or are best described by some "other" outcome. In the current data this must be inferred by comparing reported grades in the two years, but the large number of observations for which this year's grade is either less than last year's grade, or more than one year beyond last year's grade, seems erroneous. This casts some doubt on the accuracy of repeat rate estimates derived from the current data.

Returning to the same Enumeration Areas. Even with more complete information on local school supply conditions, it will be difficult to draw good inferences about the impacts of supply characteristics on enrollment rates in standard cross section regressions. School supply characteristics are likely to be correlated with other community characteristics that are difficult to measure and include in regressions. If new waves of the WMS/HICES were administered, at least in part, on a sufficiently large sample of enumeration areas sampled previously, and if (as seems likely) school supply conditions have changed in some of the re-sampled enumeration areas, then the data from pooled waves can be treated as a "panel" of data on enumeration areas.[13] Rather than observing simply how enrollment rates are correlated with the level of school supply characteristics, researchers could observe how improvements in enrollment rates across enumeration areas are related to improvements in school supply conditions. If the improvements in school supply are less correlated with unobserved community characteristics than are the levels of school supply characteristics, this will provide a more accurate picture of true school supply impacts.

[13] Note that the proposal is to return to the same enumeration areas, but not to the same households. If we are interested in studying changes in the enrollment rates of primary school aged children, then we want to begin with a new random sample of households within each enumeration area each time the survey is conducted. This allows us to observe a random sample of 7 to 14 year old children in each wave.

References

Behrman, Jere. And James Knowles (1997), "How Strongly is Child Schooling Associated with Household Income?" *Penn Institute for Economic Research Working Paper 97-022.*

Central Statistical Authority (CSA), Federal Democratic Republic of Ethiopia (2002), Statistical Abstract 2001.

Central Statistical Authority (CSA), Federal Democratic Republic of Ethiopia (March 2001), Report on the 1999/2000 Household Income Consumption and Expenditure.

Central Statistical Authority (CSA), Federal Democratic Republic of Ethiopia (November 1999), Statistical Report on the 1999 Labour Force Survey, Statistical Bulletin 225.

Filmer, Deon and Lant Pritchett (1999), "The Effect of Household Wealth on Educatonal Attainment around the World: Evidence from 35 Countries," *Population and Development Review25(1),*p. 85-120.

Frankenberg, E. (2000), "Community and Price Data," Chapter 13 in M. Grosh and P. Glewwe, *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*, World Bank.

Manski, Charles (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies 60(3)*,July, 541-92.

Mingat, Alain. 2003. "Magnitude of Social Disparities in Primary Education in Africa: Gender, Geographical Location, and Family Income in the Context of EFA," paper presented at the *Joint World Bank-UNESCO Workshop on Investment Options in the Framework of Education For All (EFA),* Ouagadougou, Burkina Faso, June 25-27, 2003

World Bank (2004), *Education in Ethiopia: Strengthening the Foundation for Sustainable Progress,* draft report.

Welfare Monitoring Unit (WMU), Ministry of Finance and Economic Development, Federal Democratic Republic of Ethiopia (March 2002), "Poverty Profile of Ethiopia: Analysis Based on the 1999/00 HICE & WM Survey Results".

Table 1: Primary School Enrollment and Completion Rates, Ethiopia and Sub-Saharan African Countries, circa 2000

	Gross enrollment ratio Grades 1-6 (%)		Cross-sectional Grade 1 enrollment rate (%) [a]		Cross-sectional Grade 6 completion rate (%) [a,/]	
	Ethiopía	Sub-Saharan Africa	Ethiopia	Sub-Saharan Africa	Ethiopía	Sub-Saharan Africa
By gender						
Boys	79.5	84.5	104.5	76.9	52.3	47.2
Girls	64.5	72.1	81.5	66.8	37.4	36.2
Index (boys=1.0)	**0.81**	**0.84**	**0.78**	**0.87**	**0.71**	**0.77**
By locality						
Urban	115.3	103.5	120.6	88.4	83.8	61.0
Rural	54.6	70.1	44.8	65.4	7.8	28.0
Index (urban=1.0)	**0.47**	**0.68**	**0.37**	**0.74**	**0.09**	**0.46**
By consumption quintile [b/]						
Richest	100.9	106.7	110.0	89.9	66.1	68.6
Poorest	65.4	62.1	88.8	53.3	41.4	23.4
Index (richest=1.0)	**0.65**	**0.57**	**0.81**	**0.59**	**0.63**	**0.34**

Note: blanks denote not computed.

a/ Rates refer to the ratio of non-repeaters relative to the population of the corresponding age group; see chapter two of Tan(2004) for a detailed explanation on the computation of these rates.

b/ Defined by ranking households according to consumption expenditure per adult equivalent.

Source: Table 4.6 in World Bank (2004), which for Ethiopia draws on analysis of the 2000 Ethiopia Welfare Monitoring Survey and the 2000 Ethiopia Household Income and Consumption Expenditure Survey and for Sub-Saharan Africa, draws on Mingat (2003), reflecting the data for the 21 African countries.

Table 2: Description of Three Datasets

	WMS/HICES 1999/2000	LFS 1999	DHS 2000
Number of households	16,668	84,340	14,072
Number of enumeration areas	1252	2331	539
Number of households per enumeration area	12 rural 16 urban	35	27
Measurement of school enrollment	"Has (NAME) currently registered for school?" 1=yes 2=no	"Have you ever attended school before?" 1=yes, registered this year 2=yes, attended school before this year 3=no, never attended school before	"Is (NAME) currently attending school? 1=yes 2=no "During the current year, did (NAME) attend school at any time?"
Percent of households in urban areas	14	15	16
Percent of 7-14 year olds enrolled in school			
All	39	39	34
Urban	84	84	78
Rural	33	32	27
Boys	42	44	37
Girls	36	33	30
Percent of 7-14 year olds who have ever enrolled in school		42	
Mean age in years among children enrolled in grade one:			
Urban	8.5		8.7
Rural	10.7		11.1

Descriptive statistics calculated employing population weights provided by surveying organizations.
Data Sources: Welfare Monitoring Survey of 1999/2000 (WMS) merged with the Household Income and Consumption Expenditure Survey (HICES) 1999/2000, which was administered to a nationally representative sub-sample of the WMS sample, as described in CSA (2001). The Labor Force Survey of 1999 (LFS), as described in CSA (1999). The Demographic and Health Survey of 2000, as described at http://www.measuredhs.com.

Table 3: Variable Definitions and Descriptive Statistics WMS/HICES 1999/2000

Variable Name	Definition	Rural Mean	Urban Mean
Number of Observations Children ages 7-14		9532	8114
Dependent Variable Registered	Indicator=1 if currently registered for school	.36	.84
Region	Indicator =1 if region is:		
Tigray	Tigray		.08
Affar	Affar		.04
Amhara	Amhara		.19
Oromiya	Oromiya		.23
Somali	Somali		.63
Bengumz	Bengumuz		.05
Snnp	SNNP		.10
Gambel	Gambela		.04
Harari	Harari		.03
Addis	Addis Ababa		.13
Dire	Dire Dawe		.05
EA Variables:			
Litshare	Share of household heads that are literate	.23	.58
	Median distance in km. to:		
Mprimary	Primary School	3.84	0.72
Mmarket	Market	7.04	1.13
Mpost	Post Office	22.33	4.24
Household Location Variables	Distance in km. from household to nearest:		
Dprimary	Primary school	4.04	0.83
Dmarket	Market	7.12	1.31
Dpost	Post Office	22.27	4.27
Household Resources (Absolute Level)	Indicators=1 if hh is in this quintile of distribution of real consumption expenditure per adult equivalent:	.29	.22
q1	Quintile 1	.24	.18

Variable Name	Definition	Rural Mean	Urban Mean
q2	Quintile 2	.19	.16
q3	Quintile 2	.15	.17
q4	Quintile 4	.11	.26
q5	Quintile 5	5.3	5.5
Adequiv	Adult quivalent hh size		
Household Head Variables		.80	.65
hmale	Indicator=1 if head male	45.8	44.6
hage	Head's age in years	.21	.61
hlit	Indicator =1 if head is Literate		
nspouse	Number of spouses of household head		
Household Structure	Fraction of household members who are:	.20	.13
hh06	Under 7 years old	.21	.21
m15o	Male and over 15 years	.23	.29
f15o	Female and over 15 years		
Child Variables	Indicator=1 if child is/was:	.47	.52
Male	Male	.15	
Age7	Age7	.15	
age8	Age 8	.13	
age9	Age 9	.13	
age10	Age 10	.09	
age11	Age 11	.14	
age12	Age 12	.10	
age13	Age 13	.11	
age14	Age 14		
Sample: Children 7 to 14 years old.			

Table 4: Comparison of Apparent Urban and Regional Effects on Enrollment Probabilities as Controls are Added to Specifications WMS/HICES 1999/2000

Urban	0.511*(0.013)	0.525*(0.012)	0.412*(0.018)	0.354*(0.019)	0.357*(0.019)	0.336*(0.020)
Tigray		0.019(0.029)	0.061*(0.027)	0.060*(0.026)	0.078*(0.026)	0.072*(0.026)
Afar		-0.203*(0.043)	-0.151*(0.038)	-0.125*(0.038)	-0.123*(0.039)	-0.126*(0.038)
Amhara		0.030(0.019)	0.038*(0.019)	0.039*(0.018)	0.043*(0.018)	0.036(0.019)
Somali		-0.329*(0.039)	-0.275*(0.033)	-0.247*(0.032)	-0.258*(0.033)	-0.257*(0.033)
Benshan		0.115*(0.033)	0.104*(0.032)	0.122*(0.032)	0.129*(0.032)	0.141*(0.031)
Snnpr		0.020(0.023)	0.001(0.023)	-0.006(0.023)	0.004(0.023)	0.005(0.024)
Gambela		0.233*(0.040)	0.202*(0.044)	0.191*(0.047)	0.206*(0.047)	0.213*(0.044)
Harari		0.181*(0.029)	0.195*(0.027)	0.161*(0.029)	0.151*(0.029)	0.162*(0.030)
Addis		0.099*(0.027)	0.079*(0.025)	0.086*(0.023)	0.080*(0.024)	0.074*(0.025)
Dire		-0.139*(0.042)	-0.102*(0.040)	-0.129*(0.038)	-0.130*(0.038)	-0.129*(0.039)
EA controls?	No	No	Yes	Yes	Yes	Yes
Household distance controls?	No	No	No	Yes	Yes	Yes
Household Resource controls?	No	No	No	No	Yes	Yes
Other household and child controls?	No	No	No	No	No	Yes

Notes: The sample includes 17,636 children ages 7 to 14. Table reports estimated probability derivatives (standard errors) implied by probit estimates. The EA control is litshare. Other controls are as defined in Table 2.

Table 5: Comparison of Estimated Community, Household and Child Characteristics Effects On Primary School Enrollment Rates across Estimation Methods WMS/HICES 1999/2000

Number of Obs.	17,636	17,636	17,636	14,861
Distance Variables				
To primary school	-0.026*(0.004)			-0.019*(0.005)
To market	0.000(0.001)			0.000(0.003)
To post office	-0.002*(0.000)			0.003*(0.001)
Truncated Distance Variables				
To primary school		-0.032*(0.003)		
To market		-0.001(0.002)		
To post office		-0.009*(0.002)		
EA median distances				
To primary school			-0.033*(0.004)	
To market			0.001(0.002)	
To post office			-0.002*(0.001)	
Household Resource Variables				
Quintile 2	0.038*(0.013)	0.039*(0.013)	0.038*(0.014)	0.059*(0.017)
Quintile 3	0.057*(0.015)	0.057*(0.015)	0.056*(0.015)	0.103*(0.019)
Quintile 4	0.072*(0.016)	0.072*(0.016)	0.074*(0.016)	0.106*(0.020)
Quintile 5	0.072*(0.017)	0.073*(0.017)	0.075*(0.017)	0.142*(0.022)
Adult Equivalents	0.009*(0.003)	0.009*(0.003)	0.010*(0.003)	0.019*(0.004)
Other Household Variables				
Hmale	-0.045*(0.021)	-0.040(0.021)	-0.044*(0.021)	-0.039(0.024)
Hage	0.012*(0.002)	0.012*(0.002)	0.011*(0.002)	0.013*(0.003)
Hage squared	-0.000*(0.000)	-0.000*(0.000)	-0.000*(0.000)	-0.000*(0.000)
Hlit	0.118*(0.012)	0.116*(0.013)	0.119*(0.013)	0.093*(0.016)
Nspouse	0.018(0.019)	0.019(0.019)	0.021(0.019)	0.032(0.022)
hh06	-0.118*(0.047)	-0.117*(0.048)	-0.118*(0.047)	-0.126*(0.058)
m15o	-0.105*(0.052)	-0.116*(0.052)	-0.120*(0.052)	-0.179*(0.063)
f15o	0.102*(0.051)	0.097(0.051)	0.094(0.051)	-0.003(0.062)
Child Variables				
Male	0.101*(0.010)	0.102*(0.010)	0.100*(0.010)	0.129*(0.012)
Age8	0.136*(0.015)	0.136*(0.015)	0.137*(0.015)	0.148*(0.018)
Age9	0.216*(0.013)	0.217*(0.013)	0.217*(0.013)	0.251*(0.015)
Age10	0.252*(0.013)	0.251*(0.013)	0.254*(0.013)	0.293*(0.014)
Age11	0.254*(0.013)	0.252*(0.013)	0.254*(0.013)	0.290*(0.014)
Age12	0.278*(0.013)	0.277*(0.013)	0.280*(0.013)	0.323*(0.014)
Age13	0.244*(0.015)	0.244*(0.014)	0.243*(0.015)	0.290*(0.016)
Age14	0.253*(0.014)	0.254*(0.014)	0.254*(0.014)	0.289*(0.016)
Community Variable				
Litshare	0.267*(0.038)	0.251*(0.038)	0.258*(0.038)	
Community Fixed Effects?	No	No	No	Yes

The sample is children 7 to 14 years old. Table reports estimated probability derivatives (standard errors) implied by probit estimates. The first three specifications also include urban and regional indicators.

Table 6: Comparison of Estimated Community, Household and Child Effects on Primary Enrollment Rates Across Urban and Rural Regions, and Across Specifications

WMS/HICES 1999/2000

	Urban		Rural	
Number of Obs.	8107	6580	9529	8281
Distance Variables				
To primary	-0.000 (0.002)	0.011(0.006)	-0.030*(0.003)	-0.025*(0.006)
To market	0.001(0.001)	-0.002(0.002)	0.000(0.001)	0.001(0.003)
To post office	-0.002*(0.000)	0.002*(0.001)	-0.001*(0.000)	0.002(0.001)
Household Resource Variables				
	0.012(0.012)	0.030(0.016)	0.054*(0.016)	0.068*(0.020)
Quintile 2	0.029*(0.013)	0.055*(0.016)	0.067*(0.019)	0.106*(0.025)
Quintile 3	0.037*(0.013)	0.059*(0.017)	0.084*(0.021)	0.113*(0.027)
Quintile 4	0.041*(0.012)	0.080*(0.016)	0.076*(0.023)	0.151*(0.034)
Quintile 5	0.006*(0.003)	0.014*(0.004)	0.009*(0.004)	0.013*(0.005)
Adult Equivalents				
Other Household Variables				
Hmale	-0.038*(0.014)	-0.035(0.018)	-0.038(0.028)	-0.032(0.035)
Hage	0.009*(0.002)	0.011*(0.003)	0.006*(0.003)	0.009*(0.004)
Hage squared	-0.000*(0.000)	-0.000*(0.000)	-0.000*(0.000)	-0.000*(0.000)
Hlit	0.085*(0.012)	0.074*(0.016)	0.086*(0.016)	0.077*(0.020)
Nspouse	0.011(0.014)	0.015(0.018)	0.023(0.023)	0.037(0.029)
hh06	-0.172*(0.040)	-0.232*(0.053)	-0.016(0.053)	-0.001(0.069)
m15o	-0.071(0.045)	-0.156*(0.057)	-0.068(0.059)	-0.091(0.077)
f15o	0.049(0.041)	-0.017(0.055)	0.015(0.063)	-0.047(0.081)
Child Variables				
	0.039*(0.008)	0.052*(0.011)	0.106*(0.011)	0.144*(0.015)
Male	0.068*(0.010)	0.088*(0.012)	0.129*(0.022)	0.133*(0.027)
Age8	0.101*(0.008)	0.130*(0.008)	0.230*(0.022)	0.268*(0.026)
Age9	0.106*(0.008)	0.138*(0.008)	0.292*(0.023)	0.346*(0.027)
Age10	0.105*(0.008)	0.131*(0.008)	0.300*(0.025)	0.357*(0.028)
Age11	0.115*(0.008)	0.145*(0.009)	0.327*(0.023)	0.396*(0.026)
Age12	0.087*(0.010)	0.118*(0.010)	0.315*(0.026)	0.387*(0.029)
Age13	0.087*(0.009)	0.112*(0.011)	0.338*(0.024)	0.391*(0.028)
Age14				
Community Variable				
Litshare	0.151*(0.029)		0.205*(0.047)	
Community Fixed Effects?	No	Yes	No	Yes

Table reports estimated probability derivatives (standard errors) implied by probit estimates. Specifications in columns 1 and 3 also include regional indicators.

Table 7: Comparison of Estimated Community, Household and Child Effects on Primary Enrollment Rates Across Subsamples, WMS/HICES 1999/2000

Sample	Boys	Girls	Children Ages 7-8	Children Ages 13-14	2 km. or less to nearest school	More than 2 km. to nearest school
Number of Obs.	8790	8846	4626	4140	11,901	5735
Distance Variables						
To primary	-0.022*(0.005)	-0.033*(0.005)	-0.042*(0.006)	-0.024*(0.005)	-0.034*(0.007)	-0.015*(0.004)
To market	-0.001(0.002)	0.002(0.002)	0.004(0.002)	-0.002(0.002)	0.003(0.001)	-0.003(0.002)
To post office	-0.001*(0.000)	-0.002*(0.001)	-0.001(0.001)	-0.002*(0.001)	-0.002*(0.001)	-0.001*(0.001)
Household Resource Variables						
Quintile 2	0.057*(0.017)	0.021(0.019)	0.029(0.025)	0.041(0.023)	0.024(0.013)	0.049*(0.020)
Quintile 3	0.075*(0.019)	0.036(0.022)	0.045(0.027)	0.026(0.024)	0.050*(0.015)	0.053*(0.024)
Quintile 4	0.089*(0.020)	0.059*(0.022)	0.113*(.030)	0.033(0.027)	0.053*(0.015)	0.088*(0.027)
Quintile 5	0.092*(0.020)	0.050*(0.023)	0.172*(0.031)	0.001(0.027)	0.067*(0.016)	0.060*(0.029)
Adult Equivalents	0.007(0.004)	0.010*(0.004)	0.009(0.005)	0.007(0.005)	0.007*(0.003)	0.007(0.005)
Other Household Variables						
Hmale	-0.063*(0.027)	-0.025(0.028)	-0.070(0.040)	-0.029(0.032)	-0.047*(0.019)	-0.017(0.035)
Hage	0.009*(0.003)	0.014*(0.003)	0.006(0.005)	0.013*(0.004)	0.012*(0.002)	0.005(0.003)
Hage squared	-0.000*(0.000)	-0.000*(0.000)	-0.000(0.000)	-0.000*(0.000)	-0.000*(0.000)	-0.000(0.000)
Hlit	0.117*(0.016)	0.122*(0.018)	0.139*(0.021)	0.062*(0.022)	0.097*(0.013)	0.115*(0.021)
Nspouse	0.032(0.024)	0.003(0.024)	-0.008(0.034)	0.075*(0.029)	0.016(0.018)	0.022(0.030)
hh06	-0.095(0.059)	-0.128*(0.065)	-0.034(0.080)	-0.323*(0.085)	-0.149*(0.049)	-0.014(0.067)

m15o	-0.114(0.069)	-0.077(0.069)	0.045(0.098)	-0.272*(0.080)	-0.114*(0.051)	-0.060(0.078)
f15o	0.012(0.068)	0.188*(0.069)	0.189(0.097)	-0.031(0.082)	0.085(0.049)	0.013(0.081)
Child Variables						
Male			0.062*(0.016)	0.130*(0.017)	0.079*(0.010)	0.098*(0.013)
Age8	0.137*(0.019)	0.135*(0.022)	0.143*(0.017)		0.106*(0.013)	0.139*(0.031)
Age9	0.231*(0.016)	0.195*(0.021)			0.167*(0.011)	0.228*(0.032)
Age10	0.240*(0.017)	0.266*(0.020)			0.185*(0.011)	0.304*(0.032)
Age11	0.261*(0.016)	0.243*(0.021)			0.174*(0.011)	0.352*(0.035)
Age12	0.292*(0.015)	0.256*(0.020)			0.193*(0.011)	0.368*(0.032)
Age13	0.277*(0.016)	0.197*(0.023)			0.174*(0.013)	0.313*(0.035)
Age14	0.272*(0.016)	0.227*(0.021)		0.011(0.016)	0.172*(0.012)	0.361*(0.034)
Community Variable						
Litshare	0.261*(0.047)	0.273*(0.047)	0.307*(0.055)	0.279*(0.050)	0.211*(0.038)	0.214*(0.056)

Table reports estimated probability derivatives (standard errors) implied by probit estimates. Specifications also include urban and regional indicators.

Table 8: Variable Definitions and Descriptive Statistics for the LFS 1999.

Variable Name	**Definition**	**Rural Mean(Std. Dev.)**	**Urban Mean(Std. Dev.)**
Number of Obs.			28,681
Dependent Variables			
attend	Whether currently attends school	0.299	0.794
jobwork	Whether worked "job" in last 7 days	0.429	0.121
houswork	Whether did unpaid household in last 7 days	0.311	0.437
Region Variables	Whether region is:		
Tigray	Tigray	0.068	0.087
Afar	Afar	0.039	0.015
Amhara	Amhara	0.177	0.210
Oromiya	Oromiya	0.272	0.278
Somali	Somali	0.044	0.025
benshan	Benishangul-gumz	0.053	0.019
snnpr	SNNPR	0.291	0.230
gambela	Gambela	0.014	0.017
harari	Harari	0.020	0.016
addis	Addis	0.002	0.078
dire	Dire Dawa	0.020	0.025
Cluster Average Variables	Within-sample share among other household heads in the same enumeration area:	0.243	0.582
litshr	Share literate	0.012	0.199
formshr	Share with formal jobs	0.109	0.216
nojobshr	Share with no jobs		
Household Head Variables		0.816	0.659
	Whether male	45.572	44.493
	Age in years	0.232	0.577
hmale	Whether literate	0.065	0.004
hage			
hlit	Whether reports religion other than Orthodox,		
htradrel	Catholic, Protestant or	0.292	0.815

	Muslim	0.012	0.196
	Whether ever migrated	0.080	0.179
hevermig	Whether has formal job		
hform	Whether has no job		
hnojob			
		6.507	6.355
Household			
Size and	Number of members	0.201	0.151
Structure	As share of total:	0.211	0.196
hhsize	Children 0-6	0.224	0.273
	Male adults	0.004	0.050
hh06	Female adults		
m15o	With formal jobs		
f15o			
hformshr		0.518	0.485
	Whether child is male	0.156	0.128
Child	Whether child's age is:	0.132	0.120
Variables	8 years	0.136	0.130
male	9 years	0.083	0.097
	10 years	0.135	0.140
age8	11 years	0.102	0.124
age9	12 years	0.100	0.132
age10	13 years	0.057	0.070
age11	14 years		
age12	Whether fails to report	0.123	0.169
age13	mother as alive		
age14	Whether fails to report	0.012	0.026
lostmoth	father as not alive		
	Whether fails to report		
lostfath	both parents as alive		
lostboth			
The sample includes children 7 to 14 years old.			

Table 9: Child Participation Rates in School and Work by Urban/Rural Region and Gender LFS 1999

	Rural		Urban	
	Boys	**Girls**	**Boys**	**Girls**
Among all children, percent	0.371	0.250	0.848	0.813
Attending School	0.544	0.306	0.137	0.093
Working in a "job"	0.209	0.435	0.334	0.519
Doing unpaid housework				
Among all children attending school, percent	0.506	0.239	0.107	0.069
Working in a "job"	0.261	0.536	0.351	0.537
Doing unpaid housework				
Among all children not attending school, percent	0.566	0.328	0.304	0.201
Working in a "job"	0.178	0.402	0.240	0.441
Doing unpaid housework				

Sample includes children 7 to 14 years old. Population weights employed when calculating descriptive statistics. See text for definition of attendance, job and housework variables.

Table 10
Comparison of Apparent Urban and Regional Effects on Primary Enrollment Rates as Controls are Added to Specifications
LFS 1999

Urban	0.495*(0.008)	0.486*(0.008)	0.291*(0.013)	0.280*(0.014)
Tigray		0.021(0.019)	0.035(0.018)	0.023(0.019)
Afar		-0.219*(0.030)	-0.166*(0.031)	-0.166*(0.032)
Amhara		0.014(0.013)	-0.012(0.012)	-0.015(0.012)
Somali		-0.208*(0.036)	-0.207*(0.033)	-0.228*(0.032)
Benshan		0.164*(0.022)	0.131*(0.022)	0.150*(0.024)
Snnpr		-0.007(0.013)	-0.042*(0.013)	-0.030*(0.013)
Gambela		0.257*(0.034)	0.173*(0.043)	0.209*(0.043)
Harari		0.161*(0.030)	0.131*(0.031)	0.141*(0.032)
Addis		0.250*(0.020)	0.097*(0.023)	0.077*(0.024)
Dire		0.077*(0.028)	0.104*(0.025)	0.113*(0.026)
EA controls?	No	No	Yes	Yes
Household controls?	No	No	No	Yes

The sample contains xx children 7 to 14 years old. . Table reports estimated probability derivatives (standard errors) implied by probit estimates. EA and household controls as defined in Table 7.

Table 11: Comparison of Estimated Community, Household and Child Characteristic Effects on Primary Enrollment Rates Across Urban and Rural Regions, and Across Specifications LFS 1999/2000

	Urban		Rural	
Number of obs.	28,681	27,541	55,641	52,323
Household Variables:	0.111*(0.008)	0.107*(0.008)	0.084*(0.006)	0.095*(0.007)
hlit	0.056*(0.012)	0.062*(0.012)	0.141*(0.037)	0.190*(0.038)
hform	0.006(0.008)	0.016(0.008)	-0.042*(0.009)	-0.030*(0.010)
hnojob	0.010*(0.001)	0.012*(0.002)	0.008*(0.001)	0.009*(0.002)
hhsize	-0.207*(0.023)	-0.200*(0.025)	-0.067*(0.023)	-0.042(0.024)
hh06	-0.012(0.028)	-0.018(0.029)	-0.104*(0.027)	-0.092*(0.027)
m15o	0.076*(0.027)	0.056*(0.028)	0.002(0.029)	-0.009(0.030)
f15o	0.084(0.052)	0.070(0.055)	0.275*(0.112)	0.179(0.112)
hformshr				
Child Variables	-0.075*(0.015)	-0.083*(0.017)	-0.049*(0.010)	-0.059*(0.011)
	-0.048*(0.009)	-0.048*(0.009)	-0.041*(0.008)	-0.055*(0.008)
Lostmoth	-0.005(0.021)	-0.007(0.022)	-0.022(0.023)	0.001(0.027)
Lostfath	0.049*(0.005)	0.053*(0.005)	0.134*(0.005)	0.155*(0.006)
Lostboth				
Male	0.173*(0.035)		0.470*(0.043)	
EA Variables	0.143*(0.038)		0.393*(0.158)	
	0.127*(0.036)		0.141*(0.048)	
litshr				
formshr				
nojobshr	No	Yes	No	Yes
EA fixed effects?				

Table reports estimated probability derivatives (standard errors) implied by probit estimates. Specifications also include urban regional indicators, as well as hmale, hage, (hage)2, and hevermig.

Table 12: Comparison of Estimated Community, Household and Child Characteristic Effects on Primary Enrollment Rates Across Urban and Rural Areas, Boys and Girls

Dependent Variable	Urban Boys			Urban Girls		
	Attend	**Jobwork**	**Houswork**	**Attend**	**Jobwork**	**Houswork**
Number of obs.						
Household						
hlit	0.092*(0.009)	-0.032*(0.008)	0.032*(0.013)	0.131*(0.010)	-0.017*(0.006)	0.018(0.013)
htradrel	-0.134(0.077)	-0.022(0.038)	0.001(0.057)	-0.112(0.078)	-0.034(0.026)	0.042(0.082)
hform	0.045*(0.017)	-0.057*(0.012)	0.078*(0.022)	0.065*(0.015)	-0.033*(0.008)	0.036(0.021)
hnojob	0.004(0.010)	-0.050*(0.007)	0.017(0.014)	0.008(0.011)	-0.040*(0.005)	0.046*(0.013)
hhsize	0.008*(0.002)	-0.000(0.002)	-0.007*(0.003)	0.012*(0.002)	-0.002(0.001)	-0.012*(0.003)
hh06	-0.151*(0.031)	0.055*(0.026)	0.130*(0.045)	-0.254*(0.032)	0.033(0.020)	0.162*(0.045)
m15o	-0.036(0.035)	-0.006(0.028)	0.124*(0.047)	0.013(0.037)	-0.028(0.022)	0.047(0.050)
f15o	0.106*(0.034)	-0.093*(0.029)	-0.000(0.047)	0.050(0.037)	-0.054*(0.020)	-0.127*(0.045)
hformshr	0.159*(0.079)	-0.047(0.064)	-0.099(0.081)	0.042(0.067)	0.064(0.036)	-0.056(0.077)
	0.081*(0.008)	0.009(0.017)	0.072*(0.017)	0.067*(0.010)	0.029(0.016)	0.104*(0.017)
	0.119*(0.006)	0.077*(0.019)	0.152*(0.018)	0.119*(0.009)	0.048*(0.017)	0.209*(0.016)
Child	0.131*(0.006)	0.187*(0.026)	0.146*(0.019)	0.141*(0.008)	0.163*(0.022)	0.235*(0.016)
age8	0.133*(0.006)	0.240*(0.028)	0.163*(0.021)	0.163*(0.008)	0.203*(0.026)	0.266*(0.016)
age9	0.144*(0.006)	0.289*(0.025)	0.200*(0.019)	0.139*(0.009)	0.270*(0.026)	0.291*(0.016)
age10	0.126*(0.007)	0.312*(0.029)	0.189*(0.020)	0.132*(0.009)	0.335*(0.028)	0.276*(0.016)
age11	0.128*(0.007)	0.346*(0.027)	0.189*(0.020)	0.108*(0.010)	0.365*(0.028)	0.279*(0.016)
age12	-0.063*(0.019)	0.010(0.014)	-0.012(0.022)	-0.087*(0.020)	0.000(0.010)	-0.000(0.023)
age13	-0.055*(0.013)	0.028*(0.010)	-0.003(0.015)	-0.040*(0.012)	0.010(0.006)	0.012(0.014)
age14	0.009(0.026)	-0.036*(0.017)	0.026(0.039)	-0.022(0.030)	-0.005(0.015)	0.008(0.037)
lostmoth	0.142*(0.040)	-0.042(0.042)	-0.120(0.077)	0.200*(0.042)	-0.001(0.025)	-0.087(0.081)
lostfath	0.163*(0.042)	-0.073(0.044)	0.242*(0.085)	0.129*(0.048)	-0.078*(0.030)	0.120(0.087)
lostboth	0.101*(0.040)	-0.232*(0.046)	0.045(0.086)	0.150*(0.045)	-0.151*(0.027)	0.089(0.089)
EA Variables		-				
litshr						
formshr						
nojobshr						
	Rural Boys			**Rural Girls**		
Number of obs.						
Household						
hlit	0.083*(0.008)	-0.004(0.009)	0.020*(0.007)	0.081*(0.008)	-0.015(0.009)	0.025*(0.009)
htradrel	-0.088*(0.017)	0.012(0.025)	-0.019(0.016)	-0.099*(0.013)	0.028(0.022)	-0.043*(0.021)
hform	0.131*(0.045)	-0.114*(0.047)	0.010(0.035)	0.140*(0.048)	-0.079(0.042)	-0.008(0.050)
hnojob	-0.046*(0.012)	-0.041*(0.014)	0.009(0.012)	-0.038*(0.010)	-0.035*(0.013)	0.015(0.013)
hhsize	0.008*(0.002)	-0.012*(0.002)	-0.003(0.002)	0.007*(0.002)	-0.006*(0.002)	-0.010*(0.002)
hh06	-0.038(0.032)	0.325*(0.034)	0.020(0.026)	-0.090*(0.028)	0.161*(0.032)	0.163*(0.033)

m15o	-0.105*(0.037)	0.129*(0.039)	0.012(0.031)	-0.097*(0.031)	0.039(0.035)	0.053(0.037)
f15o	-0.055(0.040)	0.187*(0.043)	-0.069*(0.033)	0.059(0.034)	0.247*(0.040)	-0.229*(0.042)
hformshr	0.177 (0.138)	0.011(0.163)	0.069(0.131)	0.334*(0.152)	-0.080(0.188)	0.289(0.200)
Child						
age8	0.112*(0.012)	0.105*(0.011)	0.030*(0.009)	0.069*(0.012)	0.101*(0.013)	0.071*(0.011)
age9	0.218*(0.013)	0.185*(0.011)	0.052*(0.010)	0.148*(0.013)	0.153*(0.014)	0.160*(0.012)
age10	0.293*(0.013)	0.344*(0.010)	-0.038*(0.010)	0.194*(0.014)	0.362*(0.015)	0.031*(0.014)
age11	0.342*(0.014)	0.362*(0.010)	-0.046*(0.010)	0.248*(0.016)	0.390*(0.015)	0.050*(0.016)
age12	0.371*(0.013)	0.392*(0.010)	-0.050*(0.010)	0.238*(0.015)	0.426*(0.015)	0.050*(0.015)
age13	0.384*(0.013)	0.409*(0.009)	-0.062*(0.010)	0.239*(0.016)	0.463*(0.015)	0.050*(0.016)
age14	0.389*(0.013)	0.425*(0.009)	-0.084*(0.009)	0.209*(0.016)	0.507*(0.014)	0.010(0.017)
lostmoth	-0.043*(0.015)	0.026(0.016)	-0.022(0.012)	-0.055*(0.012)	-0.003(0.015)	0.025(0.015)
lostfath	-0.053*(0.011)	0.032*(0.013)	-0.011(0.010)	-0.030*(0.010)	0.006(0.012)	0.001(0.012)
lostboth	-0.030(0.032)	0.011(0.036)	0.012(0.028)	-0.010(0.031)	-0.022(0.033)	0.001(0.036)
EA Variables						
litshr	0.473*(0.053)	-0.114(0.061)	0.090*(0.044)	0.456*(0.040)	-0.157*(0.051)	0.129*(0.054)
formshr	0.493*(0.151)	-0.828*(0.191)	0.392*(0.132)	0.277(0.184)	-0.459*(0.220)	0.258(0.174)
nojobshr	0.091(0.056)	-0.458*(0.077)	0.090(0.057)	0.188*(0.049)	-0.560*(0.077)	0.145*(0.070)

Table reports estimated probability derivatives (standard errors) implied by probit estimates. Specifications also include urban regional indicators, as well as hmale, hage, (hage)2, and hevermig.

Table 13: Variable Definitions and Descriptive Statistics
DHS 2000

		Urban	**Rural**
Number of Observations		2282	8733
Dependent Variable			
Registered	Whether registered for school	0.814	0.273
	Whether region is:		
Regions	Tigray		
Tigray	Affar	0.122	0.063
Affar	Oromiya	0.013	0.011
Amhara	Somali	0.172	0.288
Somali	Ben…	0.037	0.013
Benshan	SNNPR	0.010	0.011
Snnpr	Gambela	0.105	0.227
Gambela	Harari	0.005	0.002
Harari	Addis	0.009	0.001
Addis	Dire Dawa	0.203	0.000
Dire		0.022	0.002
EA Variable			
Litshare		0.686	0.344
Asset Variables			
Nrooms	No. of sleeping rooms	1.620	1.293
Earthfloor	Whether floor not of earth or dung	0.317	0.022
Electricity	Whether has electricity	0.773	0.004
Toilet	Whether has toilet	0.691	0.082
Land	Whether has land	0.222	0.958
Animals	Whether has animals	0.389	0.888
Cash	Whether has cash crop	0.038	0.330
Household Structure			
Hhsize	Number of members	6.691	7.111
Hh06	Children 0-6 as share of household	0.161	0.222
M15o	Male adults as share of household	0.202	0.219
F150	Female adults as share of household	0.292	0.218
	Whether mother's language is:		
Language Variables			
Amarigna	Amarigna	0.512	0.306
Oromigna	Oromigna	0.247	0.355
Tigrigna	Tigrigna	0.129	0.063
Parent Variables			
Nofather	Whether has no father	0.154	0.082
Mage	Mother's age in years	35.659	36.601

Fschool	Whether father completed grade 1	0.440	0.173
Mschool	Whether mother completed grade 1	0.473	0.074
Mbornrur	Whether mother born in rural area	0.545	0.980
Poly	Whether husband of mother has more than one wife	0.056	0.156
Notfirst	Whether woman is wife of rank second or higher	0.030	0.075
Radio	Whether mother listens to radio	0.700	0.185
BeatOK	Whether mother believes beating is justified for going out without husband's permission	0.407	0.630
Child Variables			
Male	Whether child is male	0.515	0.516
	Whether child's age is:		
Age8	8 years	0.164	0.152
Age9	9 years	0.122	0.147
Age10	10 years	0.140	0.124
Age11	11 years	0.110	0.097
Age12	12 year	0.110	0.119
Age13	13 years	0.107	0.106
Age14	14 years	0.094	0.089
	Whether among children of mother child is:		
Oldest	Oldest	0.236	0.211
Moldest	Oldest male	0.336	0.338
Youngest	Youngest	0.221	0.107
Myoungest	Youngest male	0.407	0.340

Table 14: Comparison of Apparent Urban and Regional Effects on Primary Enrollment Rates
As Controls are Added to Specifications
DHS 2000

Urban	0.558*(0.023)	0.512*(0.031)	0.285*(0.044)	0.083(0.060)	0.075 (0.061)
Tigray		-0.032(0.041)	0.045(0.041)	0.043(0.040)	-0.100(0.099)
Affar		-0.105*(0.050)	0.014(0.046)	0.017(0.054)	0.086(0.068)
Amhara		0.045(0.037)	0.051(0.033)	0.086*(0.035)	0.089(0.048)
Somali		-0.297*(0.038)	-0.189*(0.045)	-0.198*(0.041)	-0.187*(0.049)
Benshan		0.060(0.051)	0.080(0.047)	0.094*(0.042)	0.152*(0.048)
Snnpr		0.058(0.038)	0.048(0.034)	0.031(0.033)	0.076(0.051)
Gambela		0.223*(0.055)	0.206*(0.059)	0.210*(0.062)	0.260*(0.067)
Harari		0.268*(0.044)	0.336*(0.037)	0.283*(0.044)	0.285*(0.044)
Addis		0.238*(0.069)	0.208*(0.070)	0.027(0.068)	-0.026(0.063)
Dire		0.004(0.050)	0.082(0.051)	-0.010(0.061)	-0.009(0.066)
EA control?	No	No	Yes	Yes	Yes
Asset Controls?	No	No	No	Yes	Yes
Other household and child controls?	No	No	No	No	Yes

Notes: The sample includes 11,015 children ages 7 to 14 with both parents living. Table reports estimated probability derivatives (standard errors) implied by probit estimates. The EA, asset and other household controls are as defined in Table 12.

Table 15: Comparison of Estimated Community, Household and Child Characteristics Effects on Primary Enrollment Rates Across Urban and Rural Areas, and Across Specifications DHS 2000

	Urban		Rural	
Number of Obs.	2282	1556	8733	7501
Asset	0.035*(0.013)	0.064*(0.029)	0.037*(0.012)	0.042*(0.015)
Variables	0.042*(0.020)	0.094*(0.038)	0.294*(0.068)	0.221*(0.081)
Nrooms	0.117*(0.035)	0.333*(0.092)	0.262*(0.092)	0.281(0.147)
Earthfloor	0.051*(0.020)	0.091*(0.043)	0.042(0.029)	0.067*(0.034)
Electricity	-0.007(0.019)	-0.014(0.040)	0.006(0.035)	-0.011(0.033)
Toilet	0.014(0.016)	0.028(0.036)	0.022(0.020)	0.079*(0.021)
Land	0.039(0.022)	0.079(0.046)	-0.009(0.018)	0.013(0.021)
Animals				
Cash				
	-0.003(0.003)	0.002(0.008)	0.001(0.004)	-0.001(0.004)
Household	-0.057(0.072)	-0.034(0.151)	-0.068(0.060)	-0.050(0.071)
Structure	0.073(0.072)	-0.054(0.141)	-0.125*(0.063)	-0.201*(0.082)
Hhsize	0.018(0.077)	-0.008(0.165)	0.172*(0.068)	0.071(0.087)
Hh06				
M15o				
F150	0.053*(0.021)	0.051(0.044)	0.070(0.041)	0.243*(0.068)
	-0.018(0.019)	-0.048(0.047)	0.047(0.041)	0.078(0.069)
Language	0.072*(0.013)	0.143*(0.025)	0.011(0.063)	0.282(0.160)
Variables				
Amarigna				
Oromigna	-0.024(0.022)	-0.051(0.051)	-0.025(0.019)	-0.052*(0.024)
Tigrigna	-0.001(0.012)	0.011(0.024)	0.005(0.010)	-0.010(0.012)
	0.000(0.000)	-0.000(0.000)	-0.000(0.000)	0.000(0.000)
Parent	0.053*(0.015)	0.102*(0.030)	0.110*(0.021)	0.114*(0.023)
Variables	0.015(0.017)	0.003(0.037)	0.124*(0.031)	0.093*(0.032)
Nofather	-0.013(0.014)	-0.041(0.029)	-0.191*(0.044)	-0.154*(0.043)
Mage	0.010(0.032)	0.001(0.091)	0.006(0.021)	0.010(0.025)
Mage	0.034(0.023)	0.071(0.048)	-0.015(0.026)	-0.045(0.030)
squared	0.076*(0.020)	0.151*(0.040)	0.057*(0.015)	0.065*(0.020)
Fschool	-0.011(0.015)	0.000(0.030)	-0.027*(0.013)	-0.038*(0.014)
Mschool				
Mbornrur				

Poly	0.008(0.013)	0.032(0.026)	0.084*(0.012)	0.123*(0.015)
Notfirst	0.063*(0.011)	0.139*(0.018)	0.136*(0.029)	0.163*(0.031)
Radio	0.079*(0.010)	0.148*(0.013)	0.234*(0.031)	0.283*(0.031)
BeatOK	0.084*(0.011)	0.173*(0.012)	0.333*(0.031)	0.409*(0.032)
	0.087*(0.012)	0.176*(0.012)	0.413*(0.031)	0.489*(0.031)
Child	0.083*(0.012)	0.166*(0.012)	0.464*(0.030)	0.532*(0.029)
Variables	0.093*(0.012)	0.161*(0.009)	0.464*(0.031)	0.541*(0.030)
Male	0.068*(0.010)	0.137*(0.012)	0.443*(0.034)	0.524*(0.034)
Age8	0.003(0.019)	0.011(0.033)	0.001(0.016)	-0.007(0.020)
Age9	0.004(0.014)	0.019(0.024)	0.011(0.012)	0.008(0.014)
Age10	-0.006(0.018)	0.018(0.030)	0.059*(0.023)	0.090*(0.027)
Age11	0.008(0.014)	0.002(0.027)	0.007(0.012)	0.007(0.014)
Age12				
Age13				
Age14	0.082*(0.033)		0.444*(0.063)	
Oldest				
Moldest	No	Yes	No	Yes
Youngest				
Myoungest				
EA Variable				
Litshare				
EA Fixed Effects?				

Table reports estimated probability derivatives (standard errors) implied by probit estimates. Specifications in columns 1 and 3 also include regional indicators.