Africa Region Human Development
Working Paper Series

The Determinants of Schooling Investments Among Primary School Aged Children in Ethiopia

Background paper for the
2004 Ethiopia Education
Country Status Report

Julie Anderson Schaffner
The Fletcher School of Law
and Diplomacy
Tufts University

Africa Region
The World Bank

Cover design by Word Express
Typography by Word Design, Inc.
Cover photo by Julia Anderson Schaffner

Contents

Foreword

At the April 2000 Dakar World Education Forum 164 countries declared their commitment to the goal of Education for All. That commitment is echoed in the September 2000 United Nations declaration on the Millennium Development Goals (MDG) which include as one of the explicit time-bound targets, the universalization of primary school completion by 2015. These developments signal unambiguous recognition by countries as well as the international development community that education plays a central role in fostering social and economic progress. Hearteningly, they have led to more deliberate efforts in recent years to translate rhetoric into action: governments for their part are now benchmarking their education sector development plans against the goal of universalizing primary school completion, and including these plans as a key dimension of their approach to poverty reduction; while donor countries for their part, have sought to increase aid flows and direct more of the resources to support such plans.

The salutary trends notwithstanding, low-income countries, particularly those in Africa, face an uphill battle to universalize primary school completion. In numerous countries, still too many children do not start school, let alone complete the five or six years of primary schooling typically needed to become permanently literate and numerate. The situation in Ethiopia is no exception. While the country has made tremendous progress since new policies were put in place in 1994—policies that have led to more than a tripling of the gross enrollment ratio, from 20 percent in 1993–94 to 62 percent in 2001–02—the road ahead remains arduous. Only 60 percent of the children in each age cohort currently enter Grade 1, and only 60 percent among the entrants attain Grade 4, implying a Grade 4 completion rate of merely 36 percent. As in other low-income countries, raising these rates will pose significant challenges because it will inevitably require reaching out to children living in increasingly remote areas of the country, and overcoming what are likely to be rising barriers posed by costs, poverty and parental and community misconceptions about and underestimation of the value of schooling. Clearly, even as the government builds on its strong record of success since 1994, it will be important to identify potentially promising interventions to reduce the impact of these constraints on school participation.

Analytical work on the subject is particularly timely at this juncture because the government is currently in the process of preparing its Third Education Sector Development Program (ESDP3). To help strengthen the analytical underpinnings of ESDP3, the World Bank and Ethiopia's Ministry of Education prepared a stock-taking country status report (CSR) on education—to be published shortly under the title of "Education in Ethiopia: Strengthening the Foundation for Sustainable Progress." This paper by Julie Schaffner on the determinants of schooling among primary school-aged children is part of the background work for the CSR. It exemplifies the good practice of taking advantage of available household data, matched with school census data, to weave a nuanced picture of the factors that discourage school participation. While the findings are not surprising in confirming casual observation about the role of distance, income, opportunity costs, perceived values, gender and cultural circumstances, they offer a degree of specificity that is highly relevant for policy design. The paper's multivariate analysis shows, for example, that after controlling for socioeconomic differences across households, each additional kilometer of distance from the nearest primary school reduces school registration rates by two to three percentage points, up to distances of 12 to 15 kilometers. They imply that bringing schools into the neighborhoods where children live rather than situating them 10 kilometers away could boost school participation rates by as much as 20–30 percentage points—a substantial increase by any measure, and surely one that justifies emphasizing school accessibility as a central pillar in any strategy to universalize primary schooling. Other pertinent insights from the paper include the finding that the quality of schooling services matters, as does income, though the impact of the latter factor is smaller than expected and becomes significant only in specific contexts; that for many children school attendance can be compatible with child labor, though conflicts between school schedules and child labor requirements do keep some children out of school; and that gender and orphanhood matter, with boys' school participation rates exceeding girls' by an average of between 3 and 15 percentage points, and nonoprhans' participation rates exceeding that of orphans' by an average of between 5 and 10 percentage points.

The publication of this paper is intended not only to disseminate the interesting results for Ethiopia, but also to draw attention to its highly practical approach to data analysis. It is my hope that the paper would inspire similar analyses in other settings, particularly in Africa, so that policy development can be undergirded by as solid a technical foundation as possible.

Birger Fredriksen
Senior Education Advisor
Africa Region
World Bank

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 knowledgeably 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 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.

Acknowledgements

This report was prepared as background analysis for the Ethiopia Education Country Status Report, which will be published under the title of " Education in Ethiopia: Strengthening the Foundation for Sustainable Progress". The author is grateful to helpful conversations with many colleagues, including Andrew Dabalen, Stefan Dercon, Anjini Kochar and Jee-Peng Tan, as well as assistance in data matters from Susan Banki, Joanne Lim, and the Africa Household Survey Databank. Funding for this work is gratefully received from the Norwegian Education Trust Fund (NETF).

Africa Region
Africa Region Human Development
Working Paper Series

The Determinants of Schooling Investments Among Primary School Aged Children in Ethiopia

CHAPTER 1

Introduction

Since the early 1990s, Ethiopia has made great strides in expanding the reach of its education system, but many children continue to receive little or no formal schooling. Careful analysis of socio-economic disparities in schooling investments, and of the individual, household and community "determinants" of schooling investments, can be valuable inputs to discussion as the Ethiopian government seeks to set priorities and design policies that will bring more children into the formal schooling system. Fortunately, such empirical study is feasible for Ethiopia, where a large number of large household surveys have been carried out in recent years. This paper examines the five household survey datasets described in Part 2, with an eye to drawing policy-relevant inferences about patterns of schooling investment. The focus is on the schooling of children of official primary school age—7 to 14 years old.

The first question addressed is descriptive: In which socioeconomic groups are children least likely to enter or remain in school? This question may be answered simply by constructing tables that describe how schooling patterns vary across groups of interest to policy makers. The answers, presented in Part 3, are useful as policymakers seek to set priorities.

The second question probes deeper: What are the determinants of child schooling outcomes? "Determinants" are characteristics of children, or of their households or communities, that cause variation in their parents' assessment of the costs and benefits of schooling, and thus in school attendance and attainment. "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, household needs for child labor that conflict with schooling, and parental awareness and appreciation of the benefits of education. Answering this structural question is useful for shedding light on the relative importance of such policy efforts as building schools, offering scholarships, modifying school schedules to conflict less with child work, or promoting education to reluctant parents, as Ethiopia pushes forward in its efforts to increase schooling investments.

Identifying the determinants of schooling outcomes requires the careful application of econometric methods to the analysis of survey data. This, in turn, requires development of an analytical framework that guides the selection of econometric models and estimation methods, and the interpretation of econometric results. Part 4 is thus devoted to the development of the analytical framework and its general implications for econometric analysis of the determinants of child schooling outcomes. The detailed application of these methods to three datasets is described in Appendices A, B and C. Part 5 synthesizes and summarizes the policy-relevant implications of those analyses, while Part 6 offers several practical suggestions for future data collection efforts arising out of difficulties encountered during the research presented here.

CHAPTER 2

The Data

Data Sources. Schooling data from five nationally representative household surveys are employed in this paper:

- Welfare Monitoring Surveys (WMS) of 1995, 1998 and 2000 (merged in 1995 and 2000 with the Household Income, Consumption and Expenditure Surveys of the same years),

- Labor Force Survey (LFS) of 1999 (merged with school census data), and

- Demographic and Health Survey (DHS) of 2000.

All employ similar stratified random sample designs, though the sizes vary, as described in Table 2.1. Each dataset has distinctive features that render it useful for the present report.

The WMS datasets of 1995, 1998 and 2000 employ nearly identical questionnaires, allowing careful study of changes over the late 1990s. The WMSs of 1995 and 2000 were also administered jointly with the Household Income, Consumption and Expenditure Surveys (HICES) of those years (to the same households), allowing the merging of the WMS data with detailed consumption expenditure data. In addition to containing detailed information about school registration, they contain information on distances to nearest primary school and other social services.

The LFS is larger than the other datasets. While it contains less complete information about children's recent experience in school, it has the advantages of containing information on child participation in work and housework, and of containing sufficient identification of geographic areas that it can be linked to school census data at the woreda level.

The DHS employs a smaller sample, but contains richer information about the child's parents and siblings, including questions about native language and attitudes. It also contains a simple objective measure of literacy for women and men between 15 and 49 years old (in addition to a subjective measure of literacy, which is available for all individuals at least 5 years old.).

As none of the datasets is self-weighting, weights must be employed for producing all population descriptive statistics below. Conditional on a household responding to the survey at all, response rates on most questions employed here are very high. Thus I have not adjusted the weights for non-response. It should be noted that the sample frame for all

Table 2.1. Structural Characteristics and Aggregate Schooling Outcomes in Five Datasets[1]

	WMS 1995	WMS 1998	WMS 2000	LFS 1999	DHS 2000
Number of households	11,935	45,123	25,917	81,340	14,072
Number of children ages 7–14	14,297	49,327	28,822	84,508	15,220
Number of enumeration areas	929	1808	1982	2331	539
Number of households selected per enumeration area	12 rural 15 urban	25	12 rural 16 urban	35	27
Schooling outcomes					
Percent of children ages 7–14:					
Currently registered	23.7	32.6	39.3	38.9	33.6
Currently attending	–	–	–	–	32.8
Registered last year	18.1	25.6	31.7	37.2	26.0
Ever completed grade 1		–	–	–	23.8
Ever attended school2	25.8	35.0	42.0	41.7	35.4
Literate (subjective)	25.8	19.5	26.2	24.2	27.8
Percent of currently registered children ages 7–14:					
In government school	88.9	91.4	92.4	–	–
In community school	3.3	2.6	2.6	–	–
Among children registered for first grade:					
Mean age in years	10.4	10.6	10.4	–	10.7
Percent of children ages 7–14 registered last year who:					
Dropped Out	6.5	6.1	5.9	–	5.0
Repeating	17.7	18.6	14.5	–	9.3
Advanced	75.8	75.3	79.7	–	85.7

[1] Descriptive statistics in these and subsequent text tables are calculated employing population weights provided by the survey organizations. See text for full titles of surveys.

[2] See text for a description of how imperfect measures of this were constructed.

of these datasets collected by the Central Statistical Authority of Ethiopia excluded the non-sedentary populations, which are concentrated in two regions (Affar and Somali). For details on sample designs and the fielding of the surveys, see CSA(1996), CSA(1999), CSA (2001), CSA(2002) and the DHS documentation at www.measuredhs.com.

Aggregate Schooling Outcomes. Table 2.1 presents results at the aggregate level for the variety of schooling outcomes that can be examined in any of the five datasets. The first two outcomes are the percentages of children ages 7 to 14 who are currently registered for school and who are currently attending school. The number currently attending is smaller than the number of those registered, to the extent that some have dropped out during the current school year. The three WMS datasets allow measurement only of percent registered, on the basis of the direct question: "Has (NAME) currently registered for school?" The LFS begins by asking: "Are you attending school this academic year?", and then follows up with the question: "Have you ever attended school before?", to which the responses are (1) "yes, registered this year", (2) "yes, attended school before this year" and (3) "no, never attended school before". The DHS begins with: "Is (NAME) currently attending school?" and follows with "During the current school year, did (NAME) attend school at any time?" I construct the percent "registered" by adding together those who are currently attending and those who attended at any time this year. The three WMS datasets indicate a fairly rapid rise in registration over the late 1990s. The LFS gives a somewhat higher measure, and the DHS a somewhat lower measure, but they are broadly consistent. All datasets suggest that under 40 percent of primary school aged children are in school.

The indicator of whether the child attended school last year is of interest primarily because the collection of data about school attendance last year allows analysis of advancement, repeat and dropout rates. The rates of registration in the previous year in the three WMS datasets and the DHS dataset are consistent with the rising rates over the period.

All five questionnaires contain questions about the highest grade the individual completed in school. Unfortunately, the responses are not very useful, because the questions were not administered to all respondents. The questionnaires for the WMS of 1998 and 2000 and the LFS explicitly instruct interviewers only to administer the question to individuals who respond that they are able to read and write a simple sentence. While there is no explicit skip code of this sort in the WMS of 1995, examination of the data suggests that most interviewers forgot to administer the question if the respondent had not attended school last year (because the question followed several other questions that were skipped if the individual was not in school last year).

It would be useful to have direct measures of whether an individual has ever attended school. A direct measure of this is available only in the LFS, which asks children who are not currently attending school whether they have ever attended school. For the other datasets, the measures of "ever attendance" are imperfect. For the WMS datasets, the approximate measure of the share of children who have ever attended school is constructed by adding together three components: children who are currently registered for school, children who registered for school last year (but are not registered this year), and children who are reported as being literate (but who were not registered for school either last year or this year). This measure may overstate the true rate to the extent that some children who have received no schooling report themselves to be literate. (Analysis of adult data in the DHS suggests that some such reports do happen.) On the other hand, our measure may understate the true rate to the extent that some people who attended school prior to last year did not become literate.

For the DHS dataset, the approximate measure of the share of children who have ever attended school is constructed in a similar way, adding together children who are currently registered for school, children who were registered for school last year, and children who are reported as having completed grade 1 (but who were not registered for school either this year or last year). This measure may understate the true percentage of children who have ever attended school, to the extent that some children attended and left school prior to the previous year, without having completed first grade.

The main observation regarding "ever attendance" is that the vast majority of children who have ever attended school are currently in school. Thus the imperfections in the other components of the ever attendance measure for the WMS and DHS datasets are unlikely to lead to poor inferences regarding broad patterns in ever attendance.

All datasets contain subjective measures of literacy. The WMS datasets ask: "Can {NAME} read or write a simple sentence?" The LFS asks "Can you read and write?" and the DHS asks: "Is {NAME) able to read and write a simple sentence?" Despite the use of the first person in the phrasing of the LFS question, the documentation indicates that the information was collected from the household head. Subsidiary analysis of the DHS dataset suggests that these subjective literacy measures must be

Table 2.2. School Enrollment and Mean Grade by Age

	WMS 1995		WMS 2000	
Age	Percent Registered	Mean Grade (Std. Dev.)	Percent Registered	Mean Grade (Std. Dev.)
7 years	11.8	1.3 (0.6)	20.2	1.3 (1.0)
8 years	15.5	1.7 (1.0)	30.3	1.7 (1.0)
9 years	21.9	1.9 (1.2)	36.6	2.0 (1.2)
10 years	26.0	2.3 (1.4)	43.8	2.4 (1.5)
11 years	29.1	2.8 (1.8)	47.2	2.9 (1.7)
12 years	30.2	3.1 (2.0)	48.4	3.3 (1.9)
13 years	30.7	3.7 (2.3)	48.6	3.7 (2.1)
14 years	30.5	4.2 (2.7)	47.1	4.3 (2.4)

interpreted with caution. For the women and men over 15 years old who were the primary focus of the DHS, the subjective literacy measures were complemented by an objective measure, in which respondents were given a card on which was written a sentence in their native language, and were asked to read as much of it as possible. For them literacy rates were lower by the objective measure than by the subjective measure, and the differences followed an interesting pattern. Among individuals who had attended just one or two years of schooling, the gap between subjective and objective rates was especially large. The gap narrows as years of schooling increases, but does not disappear until about seventh grade. Thus individuals with just a few years of schooling (a great majority of schooled people in Ethiopia) tend to think they are literate more frequently than they are found to be literate by a simple objective test (Schaffner, 2002b).

The WMS datasets collect information on whether the school currently attended is government, private–religious, private–NGO, and other (and some additional distinctions in some years). The aggregate results in Table 2.1 indicate that the large expansion in schooling over the late 1990s was largely in government schools. The share of schooling taking place in government schools increased. Despite a policy effort to encourage community schools, their share remained small.

A striking feature of schooling data for Ethiopia is the late age at which many children start school. Many first graders are 15 or more years old. As indicated in Table 2.1, the mean age of first graders is over 10 years, despite the official start age of seven years. This complicates the interpretation of data on registration rates. Just because a child in the 7–14 bracket is not in school does not mean that the child has already quit school or will never enter school. As illustrated in Table 2.2, rates of registration rise with age until 12 or 13 years of age.

Among children who registered for school last year, dropouts may be identified on the basis of whether or not they are registered for school this year. Among those who were registered both last year and this year, repeaters may be identified on the basis of whether the grade they are registered for this year is the same or one higher than the grade they registered for last year. Unfortunately, depending on the dataset, from 1 to 5 percent of observations for 7–14 year olds have either a reported difference in grades between this year and last year that is less than 0 or greater than 1. In Table 2.1 the dropout rate is calculated on the basis of the full sample. The repeat and advancement rates are calculated in two steps. First, I take only observations for non-dropouts whose grade this year is reported to be 0 or 1 year higher than in the previous year, and calculate the (weighted) shares that report 0 and 1. Then I multiply these shares by 1 minus the dropout rate. The WMS datasets contain independent information on whether or not the child "completed" the grade in the previous year. Completing is no guarantee of advancement. A large number of children who report having completed the previous year are nonetheless currently registered for the same grade as in the previous year. There is no obvious explanation for why the DHS data imply lower repeat and dropout rates than the WMS 2000 data. This casts doubt on our ability to perceive trends in repeat and dropout rates over the three WMS surveys.

CHAPTER 3

Differences in Schooling Outcomes Across Regions and Groups

The main purpose of this section is to set out broad patterns of differences in schooling outcomes across groups defined by region, per adult equivalent expenditure level, language/ethnic group, and gender.

Gross Differences in Schooling Outcomes by Region. In setting priorities in a large country, differences in outcomes by geographic region are often of primary interest. Given sample design and size, I can examine geographic differences along two dimensions. First, and by far the most important distinction empirically, is the distinction between rural and urban areas. The second distinction is between administrative regions. To understand the extent to which the data can shed reliable light on these differences, it is useful to describe the typical sample design employed by the Central Statistical Authority in Ethiopia. In their sample frame, Ethiopia is divided into 11 administrative regions (killils), eight of which are largely rural (Tigray, Afar, Amhara, Oromiya, Somali, Benshangul-Gumuz, SNNPR, and Gambela) and three of which are largely urban (Harari, Addis Ababa, and Dire Dawa). The rural area of each administrative region is either a sample domain (reporting level for which the sample should be representative with reasonable precision) or a set of domains (allowing aggregation to that level). (Six zones in Somali and two zones in Afar are excluded from the sample frame, because of the high incidence of nomadic peoples among their populations). Thus, in principle, it is reasonable to report statistics for the rural area of each administrative region. In practice, however, the sample sizes are substantial only for the larger, predominantly rural regions (Tigray, Amhara, Oromiya and SNNP). Several large urban areas (including the urban areas of Harari, Addis Ababa and Dire Dawa) are sample domains. But the remaining urban areas (which include district capitals, even when small) are divided into domains on the basis of size (without regard to administrative region). Thus, strictly speaking, the sample is not designed to give precise estimates for urban areas disaggregated by administrative region. In practice, sample sizes seem reasonable for the four larger predominantly rural regions as well as the predominantly urban administrative regions.

Despite the small sample sizes for some cells, Table 3.1 presents school registration percentages among children 7 to 14 years old by rural–urban and administrative region distinc-

Table 3.1. Among Children Aged 7 to 14 Years Old, Percent Who are Currently Registered for School, By Region

	Rural		Urban		All	
Killil	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered
WMS 1995						
Tigray	404	19.8	426	77.8	830	32.3
Afar	163	3.7	27	82.9	190	31.1
Amhara	1851	11.1	1048	78.7	2899	18.3
Oromiya	2918	15.2	1570	64.3	4488	20.6
Somali	281	3.7	33	61.4	314	7.0
Benshangul	185	28.5	18	48.8	203	29.5
SNNPR	1900	20.3	259	71.2	2159	24.1
Gambela	149	46.2	30	65.1	179	50.7
Harari	130	26.1	222	85.4	352	61.1
Addis Ababa	142	30.2	1270	86.2	1412	85.2
Dire Dawa	156	10.4	296	73.8	452	45.8
All	**8279**	**15.4**	**5199**	**74.6**	**13478**	**23.6**
WMS 1998						
Tigray	2571	26.2	575	82.8	3146	34.2
Afar	1463	7.6	179	87.9	1642	21.6
Amhara	7086	22.7	1988	82.8	9074	28.1
Oromiya	10719	23.0	2798	77.2	13517	28.7
Somali	1970	4.9	451	50.5	2421	19.0
Benshangul	2097	38.7	445	76.4	2542	41.0
SNNPR	10068	32.7	1069	77.0	11137	36.1
Gambela	512	59.3	316	79.5	828	63.1
Harari	670	35.2	368	89.4	1038	61.8
Addis Ababa	758	49.5	1850	88.7	2608	87.6
Dire Dawa	762	22.6	411	83.0	1173	58.3
All	**38676**	**25.4**	**10450**	**80.5**	**49126**	**32.6**
WMS 2000						
Tigray	1378	31.1	759	83.5	2137	39.6
Afar	865	12.5	313	78.0	1178	22.9
Amhara	3473	32.6	1579	87.0	5052	38.0
Oromiya	5210	31.8	1951	85.5	7176	37.3
Somali	812	8.3	563	45.1	1375	21.3
Benshangul	977	46.5	399	87.6	1376	50.0
SNNPR	5720	34.7	878	73.0	6598	37.3
Gambela	311	73.4	391	82.4	702	75.6
Harari	343	53.7	268	92.8	611	74.1
Addis Ababa	411	42.9	1133	90.9	1544	89.9
Dire Dawa	421	21.7	399	76.4	820	57.9
All	19921	32.6	8633	83.8	28554	39.3

Continued on next page

Table 3.1 (continued)

	Rural		Urban		All	
Killil	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered
			LFS 1999			
Tigray	3773	31.8	2508	84.4	6281	39.2
Afar	2179	12.7	430	75.6	2609	25.9
Amhara	9858	30.6	6043	84.6	15901	35.2
Oromiya	15128	31.7	7977	81.4	23144	36.5
Somali	2435	12.2	709	50.7	3144	21.0
Benshangul	2949	43.8	550	85.2	3499	46.6
SNNPR	15800	36.3	6692	75.8	22775	39.1
Gambela	777	66.2	476	80.2	1253	69.2
Harari	1131	44.4	455	92.0	1586	66.8
Addis Ababa	122	55.4	2253	93.2	2375	92.7
Dire Dawa	1119	39.4	728	86.5	1847	69.5
All	**55271**	**32.4**	**28721**	**84.0**	**84414**	**38.9**
			DHS 2000			
Tigray	1294	23.4	164	79.5	1458	35.5
Afar	852	16.0	38	77.4	890	21.3
Amhara	2032	28.3	172	79.8	2204	32.8
Oromiya	2383	25.6	192	82.3	2575	31.4
Somali	1065	4.1	149	32.9	1214	13.5
Benshangul	1059	37.1	72	70.9	1131	39.6
SNNPR	2117	29.9	127	71.2	2244	32.7
Gambela	693	57.9	82	75.7	775	61.8
Harari	388	42.1	367	89.8	755	65.9
Addis Ababa	0	.	1196	83.8	1196	83.8
Dire Dawa	318	23.6	453	73.2	771	55.4

tions. The most striking observation is the tremendous difference in school attendance between rural and urban areas, which is far greater than the differences across administrative regions among either rural or urban areas. The rural–urban differences are so great, that I will disaggregate all further comparisons across rural and urban areas throughout the rest of the paper. Rural registration rates were rising more rapidly in the late 1990s than urban rates, in all the major regions.

Additional light is shed on the urban–rural differences by examining differences in other schooling outcomes. Sample sizes are much smaller when looking at the average age of first graders, but they indicate a strong tendency for children to start school earlier in urban areas. According to the WMS 1995 (2000), the mean age of first graders is 11.1 (10.7) years in rural areas, but 8.4 (8.5) years in urban areas, and this gap is apparent in every administrative region. Furthermore, among children who were registered for school last year, the percentage who dropped out is 10.1 (7.9) in rural areas in the WMS 1995 (2000), but only 3.0 (1.4) in urban areas. Similarly, repeat rates (among those who continue in school and for whom the difference between this year's grade and last year's grade is either 0 or 1) are 23.3 (18.3) in rural areas in the WMS 1995 (2000), and 15.1 (9.5) in urban areas.

Gross Differences in Schooling Outcomes by Economic Status. At least as important in policy discussions as regional differences are differences by economic status. Table 3.2 shows how school registration rates differ across quintiles of real total expenditure per adult equivalent using the WMS 1995 and WMS 2000, separately for rural and urban areas. A household's quintile is calculated with respect to the national distribution, thus, for example, more than 20 percent of rural households are in the lowest quintile. The quintile measures for 1995 and 2000 may not be strictly comparable. The ordering of household real per adult equivalent consumption expenditure within a year requires deflation by regional price indices. The price indices for 1995 were constructed by Stefan Dercon (Dercon, 2002), while those for 2000 were calculated by the Central Statistical Authority, and the procedures employed may not have been identical.

The two most striking observations from Table 3.2 are (1) the tremendous size of the rural–urban differences relative to differences across expenditure quintiles within rural or urban areas, and (2) the shared increase in registration rates across all expenditure quintiles between 1995 and 2000.

Gross Differences in Schooling Outcomes by Language Group. For children in the DHS 2000 whose mothers are alive and included in the survey, it is possible to identify their mother's native language. Table 3.3 reports the incidence of school registration by mother's native language. There appears to be a modestly higher incidence of schooling among Amharic (Amarigna) speakers, which is unsurprising, given that much instruction has traditionally taken place in Amharic. Differences among the major languages (Amarigna, Ormigna, and Tigrigna) are small relative to the differences between the major languages, on the one hand, and the languages spoken by smaller groups (Somaligna, Afarigna, Others), on the other hand.

Gross Differences in Schooling Outcomes by Religion. One way of distinguishing some of Ethiopia's diverse cultures is to differentiate across major religions. This is possible using LFS 1999 and DHS 2000 datasets, as in Table 3.4. Differences in schooling rates among the

Table 3.2. Among Children Aged 7 to 14 Years Old, Percent Who are Currently Registered for School, By Quintile of Real Total Expenditure Per Adult Equivalent

	Rural		Urban		All	
	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered
			WMS 1995			
Quintile 1	1647	13.1	870	64.6	2517	19.3
Quintile 2	1703	15.4	758	70.8	2461	20.9
Quintile 3	1695	14.3	860	72.8	2595	20.9
Quintile 4	1678	16.5	869	77.5	2547	24.2
Quintile 5	1556	18.8	1842	81.9	3398	33.3
			WMS 2000			
Quintile 1	2854	29.1	1790	80.8	4644	40.2
Quintile 2	2386	33.5	1506	79.7	3892	42.3
Quintile 3	1837	33.4	1335	83.8	3172	43.9
Quintile 4	1515	34.9	1416	87.3	2931	47.6
Quintile 5	1097	36.2	2123	87.7	3220	56.7

Table 3.3. Among Children Aged 7 to 14 Years Old, Percent Who are Currently Registered for School, By Mother's Native Language

	Rural		Urban		All	
	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered
			DHS 2000			
Amarigna	1778	31.5	985	88.1	2763	42.1
Ormigna	2381	25.1	336	83.7	2717	29.2
Tigrigna	929	24.8	195	85.5	1124	38.9
Somaligna	692	11.4	133	35.7	825	17.4
Afarigna	475	12.9	3	60.5	478	13.2
Other	2186	27.7	206	69.6	2392	29.5
All	**8852**	**27.3**	**1955**	**83.3**	**10807**	**33.9**

three large religions (orthodox, protestant and muslim) are not large, but the rates for all three of these religion groups are much higher than for households reporting traditional religions. Whether this is merely picking up an effect of the remoteness of geographic location remains to be seen (in multivariate analysis).

Gross Differences in Schooling Outcomes by Gender. In many places around the world girls receive much less schooling than boys. Table 3.5 presents school registration rates by gender of children aged 7 to 14, using data from the 5 datasets. Differences in registration rates between boys and girls in Ethiopia are substantial, though not unusually large by world standards. They are much larger in rural than in urban areas. The same is also true for gender differences in mean age of first graders and dropout rates (details not shown). Some

Table 3.4. Among Children Aged 7 to 14 Years Old Percent Who are Currently Registered for School, By Religion

	Rural		Urban		All	
	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered
			LFS 1999			
Orthodox	22939	32.9	17738	87.0	40880	41.7
Protestant	10393	36.8	2888	81.9	13427	39.3
Catholic	383	43.4	121	79.3	517	47.6
Muslim	18470	30.5	7848	76.8	26381	35.5
Traditional	2242	15.4	45	37.9	2287	15.5
Others	830	19.4	72	87.8	902	24.5
All	**55271**	**32.4**	**28721**	**84.0**	**84414**	**38.9**
			DHS 2000			
Orthodox	3362	28.6	1174	86.6	4536	37.2
Protestant	1141	31.5	112	80.3	1253	34.2
Catholic	73	34.2	14	98.0	87	42.4
Muslim	3562	24.6	561	75.0	4123	29.9
Traditional	301	13.5	2	0.0	303	13.5
Others	23	18.1	1	1.0	24	20.6
All	8852	27.3	1955	83.3	10807	33.9

aspects of the results are a bit puzzling. For example, the gap between boys and girls rates appears to widen between 1995 and 1998 and then decline again by 2000. Even among the three datasets pertaining to 1999 or 2000, they differ on whether they imply any boy–girl differences at all in urban areas.

Orphans. A group about which policy makers must increasingly be concerned is orphans. Both the DHS and the LFS ask whether children's mothers and fathers are alive. Thus orphans can be identified as those children for whom neither mother nor father are alive. Only in the LFS are sample sizes large enough to render examination of orphan data worthwhile. As shown in Table 3.6, orphans (89 percent of whom live with relatives) do appear less likely to be in school. Note that the substantial differences between orphans and others emerge only when the figures are disaggregated across rural and urban areas, because orphans are more likely to live in urban areas (where schooling rates tend to be higher) than other children.

Table 3.5. Among Children Aged 7 to 14 Years Old, Percent Who are Currently Registered for School, By Sex

	Rural		Urban		All	
	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered
			WMS 1995			
Boys	4225	19.9	2535	74.8	6860	27.1
Girls	3954	10.6	2664	74.4	6618	19.8
All	**8279**	**15.4**	**5199**	**74.6**	**13478**	**23.6**
			WMS 1998			
Boys	20129	30.9	4992	83.6	25121	37.5
Girls	18547	19.5	5458	77.7	24005	27.6
All	**38676**	**25.4**	**10450**	**80.5**	**49126**	**21.6**
			WMS 2000			
Boys	10305	36.4	4137	84.5	14442	42.3
Girls	9616	28.5	4496	83.2	14112	36.2
All	**19921**	**32.6**	**8633**	**83.8**	**28554**	**39.2**
			LFS 1999			
Boys	28604	38.5	13932	86.0	42761	44.1
Girls	26667	26.0	14789	82.2	41653	33.5
All	**55271**	**32.4**	**28721**	**84.0**	**84414**	**38.9**
			DHS 2000			
Boys	6301	31.0	1450	80.9	7751	36.8
Girls	5900	22.8	1562	75.9	7462	30.1
All	**12201**	**27.1**	**3012**	**78.3**	**15213**	**33.6**

Table 3.6. Among Children Aged 7 to 14 Years Old, Percent Who are Currently Registered for School, By Whether or Not They Are Orphans

	Rural		Urban		All	
	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered
			LFS 1999			
Orphans	646	25.1	685	76.4	1342	37.1
Others	54625	32.5	28036	84.2	83072	38.9
All	55271	32.4	28721	84.0	84414	38.9

The Determinants of School Attendance: Conceptual Framework and Implications for Econometric Analysis

A Framework for Thinking About Schooling Decisions

The previous section examined simple correlations between schooling outcomes and socioeconomic characteristics of importance in policy discussion. This section moves beyond analysis of correlation to analysis of causation, using multivariate analysis of the determinants of school attendance. Identification of causal (or ceteris paribus) effects of individual determinants is possible only if a comprehensive list of potential determinants is controlled for in the multivariate analysis. We must thus prepare for econometric analysis by developing a conceptual framework, which will allow us to identify a comprehensive list of determinants that might explain variation in schooling outcomes across children in a cross section.

We begin developing a comprehensive analytical framework by assuming that households decide whether or not to send a child to school in a given year through a rational decision-making process, in which they behave as if they compare the benefits and costs of sending the child to an additional year of schooling.[1] To understand the implications of such parental decision-making for both policy analysis and empirical study, it is useful to expand on four topics: parental perceptions of the benefits to schooling, parental perceptions of the costs of schooling, the potential role of poverty and lack of credit in causing parents to "discount" heavily future benefits relative to current costs, and the dependence of current choices on past and future circumstances.

The Benefits of Schooling. In most cases, 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 their old age) or vicariously (as parents enjoy seeing their children happy and better off than themselves). Schooling may help children to obtain better jobs, improve their productivity and earnings on farms or in non-farm family enterprises, improve their access to services such as health care, or improve their voice in local government. Households may also place intrinsic value on education for their children, this value increasing with the household's wealth level, if education is, as we might guess, a "normal good."

Despite a general belief among academics and policy makers that education is highly valuable, parents in some circumstances may perceive the benefits of available schooling to be low if the quality or motivation of teachers is low, if adequate teaching materials are lacking, if the curriculum is irrelevant or in an unfamiliar language, if the stagnancy of agricultural production technologies means that there are few productivity-enhancing activities available for which education is helpful, if labor markets are lacking or so distorted by frictions and regulatory restrictions that better educated workers cannot hope to get better jobs, or if educated citizens have little hope of influencing community policies and institutions. In the case of Ethiopia, where many adults have never been to school, mere lack of exposure to what schooling is and can do may prevent parents from forming positive expectations about the benefits of schooling for their children.

In addition to the intrinsic benefits of education itself, the benefits of sending a child to school may include income received through school scholarship programs, or food their children receive in school feeding programs. While the intrinsic benefits arise primarily in the future, the benefits of scholarships and feeding programs arise immediately. The more immediate benefits may play an important role in counterbalancing the (immediate) costs of schooling, especially for the poor.

The Costs of Schooling. The 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 the school. The full cost of schooling may be much greater, however, as it includes also "the opportunity cost of the child's time" and "non-monetary" costs. By "opportunity cost of the child's time" is meant the loss in household income arising when, because the child attends school, the child must devote less time to activities that expand household income. These activities include not only working for a wage, but working on a family farm or in a family non-farm business, and freeing up the time of other household members to engage in such activities by taking care of younger siblings or doing housework. In the case of girls in some cultures, school attendance may also come at the cost to the family of postponing the girl's marriage and the receipt of the associated bride price. "Non-monetary" costs may include discomfort or even social ostracism associated with sending a child to a school deemed inappropriate (perhaps because of the gender or religion of the teacher) or increased risk of children falling prey to crime (including abduction of girls for marriage), exploitation or health hazards.

Even among children facing the same official school fees, then, the true costs of schooling may differ. They will be higher for children who live further from school, where rates of crime and violence are higher, where school facilities are lacking or in disrepair, and where teachers are deemed by parents to be the wrong gender or religion to appropriately teach their child. Opportunity costs may be higher where wages for child labor are higher, in families with more productive farms or businesses, and in families with more dependent children and fewer adults who can care for them. They may also be higher in regions where the typical work performed by children is more difficult to combine with school attendance, and in which the typical work performed by adult females is more difficult to combine with child care (making it more important for school age children to care for their younger siblings).

The Role of Poverty. Even if the perceived benefits and costs of sending their children to an additional year of school were the same for poor and non-poor parents alike, the poor might choose less schooling for their children because they must weigh current costs more heavily relative to future benefits. If all households could borrow against their children's future increased earnings at a fixed interest rate, this would not be so. Any household that perceived the future benefits to be great enough that they could pay back the loan with interest, once the child has completed school and begun working, would be willing and able to send their child to school. By borrowing they could cover some or all of the costs of the education, and would thus face less need of reducing current consumption in order to send the child to school. In reality, however, it is virtually impossible to borrow against children's future earnings. This implies that when children are sent to school, the higher the economic costs of schooling (that are not compensated by scholarships or feeding programs) the more the household must reduce its savings or its current consumption. The poor lack savings, and are already consuming at low levels, thus they feel these costs especially acutely, and are more likely to find that they outweigh the potential benefits of schooling their children.

A Dynamic View of Progress Through School. For children who were in school during the previous year, the decision regarding this year's schooling is a choice between continuing in school and dropping out. (For those who continue, some advance to a new grade and others repeat.) For children who were not in school the previous year but have attended before, the decision regarding schooling is a choice regarding readmission. If they have never attended school, it is a decision regarding entry.

The costs and benefits of schooling that the household perceives this year probably depend in a complicated way on the child's current age and past schooling experience. Consider first the simplest potential implications of child age for the perceived benefits and costs of schooling. The costs of sending a child to school may first decline and then rise as the child ages. The late age of entry of many Ethiopian children suggests that the costs of sending very young children to school—whether based in culture or in concern about the physical demands or dangers of walking to school—are considered high, but fall as the child ages. On the other hand, the value of the child's time in caring for siblings or income-generating work that must be forgone if the child attends school probably rises with age. At the same time, the benefits associated with a given year of schooling probably fall as a child ages (after the threshold schooling age of 6 or 7), as their ability to learn falls. Children starting at an older age may learn less from their schooling experience, and may take longer (repeating grades more often) to attain the same level of schooling.

With intrinsic benefits that fall and opportunity costs that rise as children age, the apparent high perceived costs of sending young rural Ethiopian children to school, which cause parents to delay in sending children to school, likely imply lower ultimate schooling completion. If the costs of school attendance at age 7 could be reduced, children would begin attending while the intrinsic benefits are higher and opportunity costs lower, creating a greater possibility of remaining in school at higher ages and increasing total schooling attainment.

In addition to depending on child age, the perceived benefits of schooling may also depend on how much schooling the child has already attained. For example, among children of the same age and in households with the same current resources and distance from school, those who have already attained more schooling may be more likely to attend (if, for example, their better progress renders it more likely that the household can achieve high perceived benefits associated with secondary education) or perhaps less likely to attend (if, for example, households care only about their children acquiring literacy, and literacy has already been attained). Stopping out of school in one year may also reduce the probability of attendance in the next year, both because the depreciation of skills renders the next stage of learning more difficult, and perhaps because it is psychologically difficult to return to a lower level of schooling than one's original classmates.

This dependence of current schooling choices on past schooling outcomes has an important implication for the empirical work below. It renders this period's schooling outcomes a function of past as well as present levels of household resources, distance from school and other factors. This is important, because the data contain measures at best only of the current value of household resources, which may be quite different from their past values in the highly uncertain environment of rural Ethiopia. Current distance from school may also differ from the distances that shaped a child's past schooling experience, because of major policy efforts in recent years to build schools.

How Gender and Birth Order Enter this Framework

Gender. Schooling choices may differ between boys and girls, even within families (and thus even when facing the same school availability and economic and social conditions) for many reasons, because parents find the benefits of schooling lower or the costs higher, when considering the schooling of girls relative to boys. The perceived benefits may be lower (even when the children have the same ability) for many reasons. Girls may be expected to do little work of the sort in which productivity and earnings are improved by education or may be expected to contribute less to their parents' wellbeing in later years. They may also face greater risks to personal safety or more serious social barriers (for example, if all teachers are male and instruction by males for female students is socially unacceptable). Parents may also believe that education is simply not meant for girls.

Birth Order. 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 children of the first wife. It is difficult to generalize about what to expect in Ethiopia, given the great diversity of cultures.

How Policy Enters this Analytical Framework

Policy Options. For a policy to be effective at increasing enrollment rates, it must cause parents who are not currently sending their children to school to begin sending them to school, and to increase the number of years that children stay in school once they start. In terms of the analytical framework, it must increase the perceived benefits or reduce the perceived costs of schooling, enough that the benefits come to exceed the costs. There are many ways of doing this. Building schools in previously unserved communities tends to reduce the transportation, room and board costs of attending school, and reduces the perceived risks of sending children to school, as parents become better able to keep an eye of what is happening at school. Offering scholarships and feeding programs tends to increase the benefits in a way that is especially helpful for compensating current costs. Improving teacher qualifications and motivation, providing more and better materials, improving the relevance of the curriculum and changing the language of instruction, may improve the potential for students to acquire valuable knowledge and skills in school. Opening up markets for qualified employment, and creating the potential for technological advances in agriculture, increases the value of those skills. Education promotion efforts or adult literacy programs may improve parents' perceptions of the benefits already available. Cutting fees and modifying school schedules to conflict less with children's work requirements reduces the economic cost. Hiring female teachers, improving oversight over teachers, and allowing for the development of schools meeting certain criteria valued by ethnic or religious groups may reduce the social cost of schooling some children. Creating conditions in which parents find it safe and desirable to start their children in school at an earlier age reduces the cost of sending younger children to school, while maintaining or even increasing the benefit of staying in school at higher ages.

Conditions Under Which Policies Might Fail. Even policies that increase benefits or reduce costs may "fail" in at least three ways. First, a policy may fail to get some children into school because for the families of those children, the policy does not increase benefits or reduce costs enough to render the benefits greater than the costs. Thus even policies that seem to make intuitive sense can fail to be effective. For example, building a school in a previously unserved community reduces the cost of attending school for children in that community, but the cost reduction may not be enough, if low school quality and high rates of poverty render schooling unattractive, even when the school is close by. Furthermore, some policies need to be tailored to the local environment if they are to be effective. For example, the generosity of scholarship programs must be set sufficiently high that it brings benefits above costs for many eligible households.

Second, policies may incur costs that have little corresponding benefit in increased school enrollment, if they direct resources to children who would have been schooled even in the absence of the policy. For example, directing scholarships to families that would have schooled their children anyway will not increase schooling rates, though it may improve the current consumption level of the households involved. In fact, it may even be possible to charge higher tuition or fees to some better-off households, without causing them to take their children out of school, allowing limited government education budget resources to be stretched further. Finally, public policies may increase public school enrollments, but have little net effect on educational attainment, because the students drawn into the public schools have simply switched into public schools out of schools run by other groups.

Interactive Effects of Policy and Socio-Economic Circumstances. Whether children are sent to school or not is determined not only by policy choices, but also by the economic and social circumstances in which households find themselves. Even within a community with given school infrastructure and identical fees or scholarships available to all households, schooling enrollment will vary across households. Households with lower per capita resources (as measured by income or consumption expenditure) will tend to weigh the present costs more highly relative to future benefits, and choose less schooling. Households with more pessimistic beliefs about the potential for their children to get jobs in which education is valuable, and households who expect to share less in their children's income later in life, may choose less schooling as well. Enrollment rates may also vary across households in which the language used at home is and is not the same as the language used for instruction in the local school. Parental education and cultural factors may also cause families to differ in the priority they place on schooling children and their perceptions of the appropriateness of child labor.

The interactive importance of policy and circumstance implies that individual policies will tend to make a difference for some households but not others. Thus any one policy may have benefits that are quite unequally distributed over the population. For example, building schools may help some unschooled children in previously un-served communities, but will not help unschooled children in communities with schools, and may not help the poor even in the newly served communities. Similarly, a means-tested scholarship program may help the poor in communities with schools, but will not help the non-poor or the poor in communities without accessible schools. Reaching a broad target population will require careful construction of policy packages.

The Nature and Potential Uses of Empirical Work Based on This Framework

This section draws out the implications of the analytical framework presented above for the estimation of policy-relevant regressions examining the determinants of children's registration for and progress through school. It begins by describing the main types of regressions to be estimated below, then moving on to consider the formulation of the right hand side and various technical estimation issues.

Current Registration Regressions. A first set of regressions employs the entire sample of children ages 7 to 14 years and involves a dichotomous dependent variable indicating whether or not the child is registered for school. This dependent variable is related to current values of as many of the community, household and child characteristics of relevance to schooling decisions as are available in any dataset. Because current attendance depends on both current and past (and possibly future) values of household resources and other circumstances, it is important to keep in mind that while the determinants explicitly included on the right hand side of these regressions measure current values, they are "picking up" (albeit imperfectly) the effects of lifetime resources and circumstances. They are best thought of, then, as reflecting the long-run effect of changing lifetime resources levels and circumstances.

Ever Attendance Regressions. The second type of regression also employs the entire sample of children aged 7 to 14, but examines the determinants of a dichotomous dependent variable indicating whether or not the child has ever attended school. (In most datasets it is measured imperfectly, as described above.) The regressors employed in these regressions are the same as those employed in regressions pertaining to current registration, but their role and interpretation is somewhat different. Current household resources and distance from school play a smaller intrinsic role in determining whether the child ever attended school for children who began school in the past than they play in the determination of current registration. The same is true for children who are (in their parents' opinion) past the age for starting school. In urban areas, children who do start school tend to start at age 7 or 8. Thus for them, the resource and other variables of greatest relevance to "ever attendance" are those pertaining to the years in which the child was 7 or 8, which is in the past for many children. In rural areas, children tend to start school later, thus it is less clear what the ideal date for measuring determinants would be. Either way, however, current values of resource, distance to school and other variables are standing in (imperfectly) for the values in the time periods most relevant to the school entry decision. (Most likely, the older the child, the greater the difference between the relevant measures we would like to include and the current measures we are forced to include. We return to this measurement problem below.)

Transition Regressions Conditional on Attendance Last Year. Four types of regressions are run on the sub-samples of children who reported attending school in the previous year, and employ dependent variables describing their transitions through or out of school. What the datasets reveal about their progress through school can be summarized by four mutually exclusive and exhaustive indicator variables: dropout, repeat, advance and unclear. Dropout equals 1 if the child is not registered for school in the current year (despite having been registered in the previous year). Repeat is equal to one if the child is registered in the current year and reports being registered for the same grade as in the previous year. Advanced is equal to one if the child is registered for school in the current year, and reports being in a grade that is one higher than in the previous year. Unclear is equal to one if the child is registered for school in the current year, and reports being in a grade that is neither the same as nor one higher than the grade for which the child reports being registered in the previous year. While it is clear that these students have continued in school, one suspects that there are errors in their reporting of grades, rendering it unclear whether they have advanced or repeated.

These regressions condition on children having been in school last year. For two reasons, this requires us to exercise caution in comparing coefficient estimates on particular regressors between these regressions and the regressions run on the entire sample. First, notice that by their very structure, the regressions control partially for schooling history, because attention is limited to children whose past resource levels were adequate (given their preferences, abilities and opportunities) to get them into school last year. Coefficients on current resources and circumstances will thus pick up only part of the total effect of lifetime resources on these schooling outcomes. The effects of previous resource levels that affect current choices by determining whether the child was in school last year are purged from the coefficients on household resources in such regressions. Thus we would expect, for example, to pick up less of an effect of household resources on transition rates than on ever attendance or current attendance.

Second, the samples on which these regressions are run are "endogenously selected." Regressions for these variables are based on the sample of children who registered for school last year. Whether or not a child was registered for school last year is endogenous to this year's schooling outcomes, in the sense that last year's school attendance depends on many of the same observed and unobserved characteristics and circumstances that determine this year's outcomes. For example, children who registered last year are likely to be from households characterized both by higher resource levels (an observed characteristic in some datasets) and by greater inherent interest in educating their children (an unobserved preference or taste characteristic).

Endogenous sample selection can lead to bias in the estimation of the effects of regressors on the transitions (dropouts, etc.). The intuition behind this bias is best explained by focusing on a specific example. Suppose we are interested in estimating the effect of increasing observed household resources on dropout rates among children currently enrolled in school. Suppose also (contrary to reality, for the sake of simplicity) that dropout rates depend positively on just one observed characteristic, household resources, and one unobserved characteristic, household preferences toward education. Suppose furthermore, that the same two factors (household resources and household preferences) have positive effects on whether the child was registered for school last year. Finally, suppose estimation conditions are relatively good, in the sense that household resources and household preferences are uncorrelated in the population as a whole. (Under such circumstances regressions of child schooling outcomes on the observed determinants in the entire sample should yield unbiased estimates.) Under these circumstances, if we observe that a child attended school last year but had very low household resources, then we can infer that that household must have had quite high levels of household preferences toward schooling. By contrast, children from households with very high resources may be from households with widely varying levels of preferences, and still be observed in the sample of children who attended last year. Thus, as household resource levels increase among children we know attended school last year, we can infer that, on average, household preferences toward educating children are falling (even though resources and preferences are uncorrelated in the population of children aged 7 to 14 as a whole). This is the sense in which endogenous sample selection induces an omitted variables bias.

In the case just described, endogenous sample selection suggests a downward bias (in absolute value) in the estimation of the effects of household resources on dropout rates. That is, while we would like the estimates to reflect the ceteris paribus effect on the dropout rates of increasing household resources, they will instead tend to capture the net effect of increases in household resources and of the associated reduction in average household preferences toward education that is relevant in the endogenously selected sub-sample. This suggests that if we find smaller effects of household resources on dropout rates than on rates of "ever attendance", we don't know if this difference is "real" or whether it is an artifact of the endogenous sample selection.[2]

Careful treatment of the endogenous sample selection problem is beyond the scope of the current analysis. Standard, simple applications of Heckman's two-step procedure are unlikely to produce reliable estimates, because they are based on a variety of strong assumptions that may be incorrect. More careful use of procedures in the spirit of Heckman's two-step procedure, but involving model selection testing and sensitivity analysis, have greater promise and are feasible (Schaffner, 2002a), but are beyond the scope of the current analysis. Thus the results presented below for the dependent variables dropout, repeat, advance, and unclear must be interpreted in the light of possible sample selection biases. Caution should be exercised in comparing coefficients on the same regressors across these regressions and regressions for current registration and ever attendance.

Time Use Regressions. The LFS/EMIS data (described in Appendix B) allow examination of three dependent variables describing the current use of children's time: whether the child is currently attending school, whether the child is engaged in income generating activities and whether the child participates in unpaid housework. Because child time use and schooling are determined simultaneously, the same regressors relevant for school registration regressions are potentially relevant here. We hope that regressions of these dependent variables on the same set of variables that determine children's school registration status will shed some light on the extent to which the need to devote child time to paid work, work in family enterprises or child care may prevent some children from attending school.

The Determinants of School Attendance and Indicators of Progress Through School. The framework examined above implies that whether or not a particular Ethiopian child attends school (and whether or not he or she progresses through school) will depend on a wide range of factors. Table 4.1 attempts to provide a reasonably comprehensive list of these factors. The characteristics listed under "Availability and Quality of Schooling in the Community" are all factors shaped by school supply policy, and common to all households within small geographic areas. When their effects can be estimated explicitly and well, their inclusion allows assessment of school supply policy impact, and identification of the conditions under which school supply improvements have larger and smaller effects, with direct policy implications. "Other Community Characteristics" include other economic and social conditions faced in common by households within geographic areas. Understanding their effects may shed light on the role of local perceptions regarding the benefits of education, as well as the role of labor market tightness in shaping the opportunity cost of children's time.

Table 4.1. List of Potential Determinants of Whether a Primary School Aged Child Attends School

Availability and Quality of Schooling in the Community
Numbers of primary school classrooms, sections or teachers per child in region
Numbers of secondary school classrooms, sections or teachers per child in region
Geographic dispersion of available schools
Languages in which instruction is available in region
Availability of alternative school schedules in region
Availability of schools related to specific religions or groups in region
Availability of schools with latrines and other facilities
Shares of teachers with varying levels of qualification and skill
Availability of female teachers
Numbers of textbooks per child in region
Quality and relevance of curriculum employed in region
Motivation level of teachers
Structure of school fees, requirements to purchase books, uniforms or supplies
Availability of scholarships or fee waivers, school feeding programs

Other community characteristics
Quality and quantity of roads
Quality of local law enforcement
Significance of local health hazards to which exposure is increased by attending school
Levels of wages and unemployment for children and adults
Potential to adopt new technologies or switch into new economic activities
Potential for broad-based involvement in local politics
Local customs regarding marriage
Local beliefs about value of education
Local exposure to people with education

Household Resources and Needs
Level of per adult equivalent real consumption expenditure
Household size, Number of School Aged Children

Other Household Characteristics
Distance from school
Access to local labor market
Numbers of household members in various age and sex categories
Ownership of land, livestock and other household enterprise assets
Language
Ethnicity
Religion

The third group, “Household Resources and Needs,” captures the role played by income poverty in determining schooling outcomes. What we want to capture in the empirical work is the level of income available to the household for meeting its needs, and the magnitude of the needs over which this income must be stretched. Consumption expenditure in the last month (expressed in real or deflated terms) is thought to be a good measure of a household’s annual income. In seasonal environments, where incomes fluctuate a great deal from month to month, households may find ways to save when times are good and dis-save when times are bad, in order to smooth out their consumption relative to their fluctuating income. Current income in any one period, then, might give a pretty poor indication of the level at which the household typically lived during the year, while consumption expenditure will give a more accurate indication. Mechanisms for saving and dis-saving may not be strong enough to smooth fluctuations in income from year to year induced, for example, by drought or conflict. Thus while consumption expenditure in the last month may provide a reasonably good gauge of living standards and resources in the current year, it may provide a very incomplete description of the resource levels that have shaped child schooling outcomes in the recent past.

Whatever income the household has must be stretched to cover the food, clothing and shelter needs of all the household members, as well as the possible schooling of school-aged children. The main role of household size is usually acknowledged by expressing consumption expenditure on a per capita (or per "adult equivalent") basis. If there are economies of scale in meeting basic needs out of consumption expenditure, then among households facing the same per adult equivalent consumption expenditure, larger households may be less resource constrained. Thus household size may have an independent effect on schooling outcomes, even after controlling for per capita consumption expenditures.

"Other Household Characteristics" identify additional reasons why schooling choices may differ across households, even within communities in which all households face the same availability and quality of schooling and the same general economic and social circumstances. This list includes factors describing social and economic status of the household, its geographic location within the community, the type of assets it owns, and its demographic structure, which may shape the opportunity cost of the time of school-aged children.

"Parental Characteristics" is a shorthand way of referring to the characteristics of the adults who play a primary role in deciding whether a child attends school or not. In nuclear family households, these individuals are both the child's parents and the household heads. In more complex family and household settings, there is more diversity in the right way to identify these individuals. The child's parents may not be the household heads. When not the household heads, they may nonetheless retain the primary role in making choices about their child's schooling, or they may not. However these decision-makers are identified, the characteristics listed in the table may influence schooling outcomes. "Child Characteristics" are included in the table, because even among children with identical community, household and parental characteristics, schooling outcomes may differ systematically in ways related to the factors listed there.

The list in Table 4.1 includes factors of potential importance to schooling outcomes, whether or not they are measured in the datasets available. Thinking carefully about what is excluded from a multivariate analysis is of great importance when attempting to interpret empirical patterns.

The Importance of Including a Comprehensive Set of Controls in School Attendance Regressions. Several sets of variables are included in multivariate analyses of school attendance or school registration because they are of direct policy interest. Five sets of variables stand out. First, including school availability and quality variables (if included in a well-specified regression) allows direct measurement of the impact on school attendance of specific changes policy-makers might contemplate in local school supply. Second, including measures of household resources allows assessment of the likely importance (under various circumstances) of changes in policies regarding school fees and scholarships, and of the likely impact of more general economic development, on school attendance. Third, including variables thought to influence the opportunity cost of the child's time may shed light on the nature and magnitude of these opportunity costs, and how they vary across children of different types and in different locations. Fourth, including proxies for exposure to the potential benefits of schooling and beliefs about schooling may allow assessment of the extent to which parental attitudes and beliefs may play an important role in keeping children out of school, even when schools are available and household resources permit. Fifth, including indicators of the language, gender, and other characteristics of students might help uncover troubling differentiation across groups in the extent to which they are being served by public schools.

It is important to include controls for other characteristics, even when measurement of their effects is not of direct policy interest, in order to avoid bias in the estimation of the effects of primary policy interest. As is well known, omitting relevant variables from multivariate regressions leads to bias in the estimation of the effect of a variable included in the regression, if the omitted variables are correlated with the included variable of interest (after holding constant the level of other included variables). Thus, for example, the omission of household resource controls from the regression might cause the coefficients on school supply variables to "pick up" some of the effect of household resources, and overestimate the impact of school supply, if school supply variables tend to be better in communities where households are better off. The power of multiple regression lies in its ability to tease out causal (or ceteris paribus) effects of changes in specific determinants on schooling outcomes, provided the regression contains a comprehensive set of controls for the factors that might cause variation within the sample in the schooling outcome.

Explicit Controls for Potential Schooling Determinants. Controlling explicitly for a determinant listed in Table 4.1 involves measuring the determinant directly and including it in the regression. Only when we include explicit controls (and when those controls vary significantly within our sample) do we have any hope of quantifying the effects on schooling outcomes of specific changes in community and household circumstances. In the datasets employed below, explicit controls of three sorts may be employed:

- Household survey data on child, parent and household characteristics
- Within-sample averages of surveyed characteristics within small geographic regions called enumeration areas
- School and population census information aggregated to the woreda level and matched to household survey observations in the same woredas

The "enumeration area" (EA) is the primary sampling unit in all the household surveys conducted by the Ethiopian Central Statistical Authority (CSA). They are neighborhoods of around 200 (100) households in urban (rural) areas. Sample selection for CSA household surveys proceeds in two stages. In the first stage, a stratified and representative random sample EAs is drawn. In the second stage a random sample of households is chosen from within each selected EA. The number of households chosen per EA ranges from 12 (WMS of 1995 and 2000) to 35 (LFS). Within-sample averages of characteristics by EA (for example, the share of household heads within the sample from a particular EA that are literate) provide noisy measures of community circumstances. Such measures are likely to be somewhat more reliable in the LFS than in the other surveys, given the larger sample sizes within EA s.

The woreda 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. Woreda popoulations range from about 9000 to about 535,000, with mean of about 139,000 (according to the CSA, 2002). While such geographic units are larger than the "communities" within which school supply and other community characteristics influence households' schooling choices, they are the smallest geographic units for which school census data may be matched to the household-level data of the LFS, as well as to Statistical Abstract data on population density.

Data constraints cause the methods of controlling for potential schooling determinants to vary across the three datasets employed in multivariate analysis: the WMS 2000 merged with the Household Income, Consumption and Expenditure Survey (HICES) 2000, the LFS 1999 merged with school census data (EMIS), and the DHS 2000. A rich set of explicit controls for the availability and quality of schooling in the community is available only in the LFS/EMIS 1999 data. Unfortunately, they are measured only at the woreda level, which limits their usefulness (as is discussed in Appendix B). In the WMS/HICES 2000 data, household distance to the nearest primary school captures both availability of schools within the community and the household's location within the community. The DHS contains no controls for local school supply.

Explicit controls for other community characteristics are of two types: a population density measure at the woreda level (merged with the LFS/EMIS data on the basis of data reported in the Statistical Abstract, 2001), and EA-level means (calculated over all observations other than the one for which the calculation is being made) of characteristics recorded in the WMS and LFS surveys.

Explicit controls for household resource levels are available only in the WMS/HICES 2000 data, in which it is possible to calculate real per capita adult equivalent consumption expenditure quintile classes. Resource levels are controlled for imperfectly in the DHS data through the inclusion of indicators of various assets the household may own. (As discussed in Appendix C, these asset variables seem to proxy simultaneously for household resources and "urbanicity" of location.) No explicit controls for household resource levels are included in the LFS/EMIS analysis, though indicators of the extent to which household members are involved in formal employment probably capture some important variation in levels of prosperity.

In all three datasets it is possible to control for household structure, using the percentages of household members who are in several age–gender categories. A somewhat richer depiction of household structure, language and religion is possible in the DHS data.

In the WMS/HICES and LFS/EMIS analyses, "parental" characteristics shaping preferences toward schooling for the children are proxied by characteristics of the household head, because the characteristics of the child's parents cannot be identified (with the exception that the LFS allows identification of whether the child's parents are alive). In the DHS analysis, the characteristics of the child's mother may be employed explicitly.

In all datasets it is possible to include indicators of the child's age and gender. In the DHS and LFS/EMIS it is possible to include as well indicators of whether the child has lost one or both parents. In the DHS it is furthermore possible to identify the child's birth order (among children of the same mother).

The effects of factors that cannot be controlled for explicitly, or for which the explicit controls are imperfect measures, may be picked up implicitly by regional indicators, or by the inclusion of "fixed effects" at the enumeration area or household level. It is thus useful to discuss these implicit controls.

Indicators of Rura//Urban Residence. Indicators of rural/urban location may be used to control implicitly for the effect of any variables omitted from the regression, whose means tend to differ between urban and rural areas. In a regression including a rich set of variables describing household resources and circumstances, but no explicit controls for school availability and quality, the regional indicators will tend to pick up the effect of average differences across regions in school supply (and in other community characteristics). Because they control only implicitly, their coefficients allow no direct assessment of school supply effects, but they reduce the tendency for omitted variables bias in the estimation of household resource effects. Unfortunately, they control only imperfectly for the effects of school supply and other community characteristics. School supply and other community characteristics may vary greatly across communities within rural or urban areas and within regions. If this within-region variation in correlated with typical levels of household resources, then the estimation of household resource effects remains biased (though probably less so) after the inclusion of rural/urban indicators.

As seen above, differences in schooling outcomes between rural and urban areas are more profound than differences along any other dimension observed in the data. Thus it is especially interesting to examine competing explanations for these differences. Thus the appendices examine the extent to which estimated rural–urban differences in schooling outcomes diminish as controls for school supply conditions, household resource levels and other determinants are added to the regressions.

Region Indicators. As with rural/urban indicators, indicators of major region of residence pick up the effects of school supply, household resources, attitudes and other factors that differ systematically across regions. Supply conditions are likely to vary across regions, given the federal structure of government, in which federal resources are block granted to the regions, which (at least in principle) have a high degree of autonomy.

Enumeration Area Fixed Effects. Some omitted variables may vary within broad regions, but be largely shared by households within smaller geographic areas. Including EA "fixed effects" is equivalent to including in the regression indicator variables for every EA (minus one) represented in the data. Each EA's "effect" (equivalent to the coefficient on its dummy variable) "picks up" the effect on schooling outcomes of the average level within that geographic area of all variables omitted from the regression. They probably pick up important differences in school quality and availability, as well as differences in the structure of economic and social opportunities within regions. As with the regional indicators, they are implicit controls, so their coefficients are of no direct interest in quantifying policy effects. Their inclusion, however, may be important for reducing omitted variables bias in the estimation of other effects controlled for explicitly in the regression. Unfortunately, when EA fixed effects are included, they pick up the aggregate effect of all EA-level factors. It is thus no longer possible to include in the regressions any variables that vary only from EA to EA, including the school availability and quality variables derived from school census data.

It is possible to incorporate EA fixed effects into the regressions from all the datasets examined here, because the samples are structured in such a way that multiple observations are available for each EA (as described above). To gain intuition about interpretive differences between regressions that do and to not allow for EA fixed effects, it is useful to think of EAs as "communities" of relevance to the schooling decisions of households. EA fixed effects methods essentially include in linear regressions individual dummy variables for each community contained in the sample. This embodies the assumption that there are features of communities that tend to increase or decrease schooling outcomes for all households in those communities. By including this full set of community dummies, we control for all the characteristics of communities that influence schooling outcomes, without having to identify and measure exactly what those features are. When EA fixed effects are controlled for in estimation, the coefficients on household characteristics like distance to school or consumption expenditure levels are estimated off of within-community variation in the values of those variables. This reduces the potential for coefficients on household- and child-level variables to suffer from biases associated with the omission of community level variables.

In the absence of EA fixed effects, we might worry that coefficients on household resources are biased, because households with higher resource levels tend to live in more modern communities. That is, the apparent effect of income on schooling outcomes might instead be picking up effects of "community modernity" on household preferences toward schooling. Once we include EA fixed effects, however, we implicitly control for community modernity (and other community characteristics), allowing us to derive an estimate of household resource effects that is purged of such biases.

Household Fixed Effects. Unobserved factors relating to tastes for schooling probably vary across households even within EAs, but would be shared by children in the same households or children of the same parents. Including household fixed effects would control for such factors. Such specifications allow inferences only on the effects of variables that vary across children within households, such as gender, birth order, and orphan status.

Estimation Within Sub-Samples. The analytical framework suggests many potentially important reasons why the size of the effect of one determinant on schooling outcomes may differ depending on the levels of other determinants. For example, school attendance may vary little across households with different income levels where there are no schools, while income may play an important role in explaining school attendance among households who live near to schools. A convenient way to allow for a broad range of such interaction effects is to break the sample up and estimate schooling relationships within sub-samples. It will be especially useful to examine contrasts in regression results along the following lines: rural versus urban, boys versus girls, high versus low income, households near to primary schools versus households far from primary schools, and younger versus older children (within the 7–14 range). The appendices expand on some of these distinctions.

Estimation Methods and Reporting of Results. All the dependent variables are dichotomous. The preferred method for relating dichotomous dependent variables to their potential determinants is probit regression. The assumption underlying this method is that the probability that the dependent variable equals one (e.g. the probability that a child is registered for school) is equal to $\Phi(X\beta)$, where $\Phi(.)$ is the standard normal cumulative distribution function, X is a set of regressors and β is a set of coefficients to be estimated. (By contrast, a linear regression would assume that the same probability equals $X\beta$. This assumption is unappealing for several reasons. Perhaps the most important is that it allows predicted probabilities of school enrollment that are less than 0 and greater than 1.)

While the magnitudes of the estimated probit coefficients (β) themselves can be difficult to interpret, it is straightforward to calculate so-called "probability derivatives", which are much easier to interpret. The probability derivative associated with the estimated coefficient on a particular right hand side variable is an estimate of the (approximate) effect of increasing that right hand side variable by one unit on the probability that the dependent variable equals one (e.g. the probability that the child attends school). In the probit model, this probability derivative is equal to $\phi(X\,\beta)\,\beta j$, where βj is the coefficient on the particular regressor of interest. A probability derivative equal to, say, 0.10, indicates that a one-unit increase in X is associated with a 10 percentage-point increase in the probability that the dependent variable equals one. If the dependent variable is an indicator of whether or not the child is currently registered for school, the probability derivative indicates the percentage point increase in school registration rates associated with a one-unit increase in the relevant control variable. Probability derivatives are thus transformations of the estimated coefficients into units that are easy to interpret.[3]

Given the functional form assumption of the probit model, the value of the probability derivative varies with the level of the X's, but it is customary to report the value derived by plugging in the mean values of X in the dataset, as well as the estimated coefficients. In all the tables presented below, probability derivatives evaluated at sample means are reported (rather than reporting the estimated coefficients themselves) in order to facilitate interpretation of results.

For reasons discussed below, some results are also reported for OLS rather than probit specifications. OLS is the appropriate method of estimation under the assumption that the probability that the dependent variable equals one is a linear function of the regressors. Under this assumption, the probability derivatives of interest for interpretation of results are constant for all values of the X's, and are equal to the estimated coefficients themselves. The magnitudes of the coefficients in OLS estimation are thus directly comparable to the magnitudes of the probability derivatives reported for the probit specifications.

If estimated probability derivatives are highly similar across probit and OLS specifications, then the estimated effects can be thought of as "robust to underlying functional form assumptions." This is true in many applications and increases confidence in the results. Unfortunately, as we will see below, the results tend not to be highly robust across the two specifications in the WMS data. Being based on somewhat superior functional form assumptions, the probit specification should be preferred to the OLS results.

Despite their weaknesses, the OLS results are also included, because it is desirable to introduce two complications into the estimation methodology that are much more feasible to incorporate into the linear OLS models than into the nonlinear probit models. The first is the introduction of "enumeration area fixed effects", discussed above. The second is the use of "instrumental variables" methods, which are of interest for dealing with potential measurement error biases associated with the estimation of school distance effects, as discussed in Appendix A.[4]

In all tables of regression results, reported standard error estimates are calculated using formulas that account for arbitrary heteroscedasticity, as well as the clustering of observations within enumeration areas. Such estimates are consistent in the face of problems that can render more standard calculations of standard errors misleading. Asterisks in the tables of probit and OLS regression results indicate that the coefficient estimates were significantly different from 0 at the two-tailed five percent level. In the case of the probit model, the coefficients whose significance is indicated are the underlying coefficients on the individual right hand side variables (which are not reported in the tables) rather than the probability derivatives that are derived from the estimates (a standard procedure).

Estimation results were produced using the statistical software package STATA, the commands dprobit, regress, xtreg and ivreg, and with options robust and cluster( ).

Notes

1. If more than one schooling alternative is available (e.g. public and private), they send the child to the school for which the excess of perceived benefits over costs is greater. Because the vast majority of primary schooling

is public in Ethiopia, the model here does not dwell on public/private choices.

2. Note that this is not just a problem of being unable to extrapolate results from the group of children who registered for school last year to all other children. Even restricting our attention to the select group of children who were already registered last year, this bias causes our estimates to fail to capture true ceteris paribus effects and thus may render our results misleading.

3. For right hand side variables that are dichotomous, it is customary to replace the calculation of "probability derivatives" described here by the calculation of the impact on the estimated probability of changing the value of the dichotomous regressor from zero to one, while holding all other regressors at their means. I follow this standard in my reporting of "probability derivatives."

4. It is feasible to introduce EA fixed effects into probit estimation employing statistical software with the capacity to employ large quantities of memory, such as STATA SE. The results are broadly similar to results presented below. In this paper I restrict attention to methods that are feasible with more broadly disseminated statistical packages, with less memory, such as intercooled STATA.

CHAPTER 5

The Determinants of School Attendance: A Synthesis of Econometric Results for Three Datasets

The previous section described a general approach to the econometric analysis of household survey data, for the purposes of drawing policy-relevant conclusions about the determinants of household decisions regarding child schooling. I have applied that general approach to the analysis of three datasets. Details regarding the analysis of the WMS 2000 (merged with the Household Income, Consumption and Expenditure Survey of the same year) may be found in Appendix A, while the details of the analyses of the LFS 1999 (merged with school census data at the woreda level) and of the DHS 2000 may be found in Appendices B and C. This section synthesizes the results of all three analyses, as well as additional descriptive information, in a discussion of what the empirical evidence suggests for education policy in Ethiopia.

The Role of School Supply. To the policymaker concerned with how to spend scarce resources on improving primary schooling rates, a first question has to do with the potential importance of simply building more schools, and of staffing and equipping existing schools better. Evidence on this comes from two main sources: (1) studies of the relationship between households' distance to the nearest primary school and household schooling choices, using WMS/HICES data (Appendix A); and (2) studies of the association between school supply characteristics at the woreda level and household schooling choices using the LFS/EMIS data (Appendix B).

To put the econometric results into context, it is useful to provide a brief description of school supply in Ethiopia. The three WMS datasets contain questions on the distance from the household to the nearest primary school. Table 5.1 describes the distributions of children ages 7–14 by reported distance to the nearest primary school. The general picture is one of gradual improvement in school supply, but continued serious supply constraints. Despite reductions over the late 1990s in the percentage of children living far from the nearest primary school, nearly a third of rural children continued to live at least 5 km. from the nearest primary school in 2000.

Some aspects of the results are puzzling. For example, while the table shows some decline in the percentages of rural children that are very far from the nearest primary school, we also see declines in the percentages of rural children who are less than one km. from the nearest school, and declines in the percentages of

Table 5.1. Percentage Distribution of Children Ages 7 to 14 by Distance to Nearest School

Distance to Nearest Primary School in Km.	WMS 1995	WMS 1998	WMS 2000
	Rural		
0	19.5	19.1	15.4
1–2	23.4	24.1	28.1
3–4	23.0	23.5	27.8
5–6	15.9	18.3	18.3
7–12	13.6	11.8	11.8
Over 12	4.5	4.5	3.2
	Urban		
0	67.9	56.7	47.4
1–2	30.2	36.1	45.3
3–4	1.3	6.4	6.2
5–6	0.2	0.7	1.0
7–12	0.3	0.0	0.1
Over 12	0.1	0.1	0.0
	All		
0	26.1	24.0	19.6
1–2	24.3	25.7	30.4
3–4	20.1	21.3	24.9
5–6	13.8	16.0	15.0
7–12	11.8	10.3	8.3
Over 12	3.9	2.8	1.8

urban children who are less than a km. and between 1 and 2 kms. away from the nearest primary school. If these changes were "real," they would indicate some tendency toward school consolidation that reduced the spread of schools within regions that were already fairly well covered by schools. It seems more likely, however, that they are the artifact of changes in the nature of the sample or in the nature of errors in the reporting of the distances by respondents.

Table 5.2 presents estimates based on the three WMS datasets. While registration rates decline as distance to school increases, in both rural and urban areas, the most striking observation from the table is that differences in distance to school play a remarkably small role in explaining differences in schooling rates between rural and urban areas. Even when attention is restricted to children in rural and urban areas who are within a kilometer of the nearest school, registration rates in rural areas are much lower than registration rates in urban areas, though they are rising more rapidly over the late 1990s in rural areas than in urban areas. A related observation is that much of the increase in school attendance rates in rural areas over the late 1990s appears to be taking place within sub-groups of children living a given distance from school. Finally, notice that even among rural children living less than a kilometer from school, fewer than half are registered for school. This suggests strongly that availability of schools is not the only constraint on attendance, nor is the expansion of school availability the only force behind rising registration rates.

Calculations of the mean age of first graders by distance to nearest primary school tell a similar story. According to the WMS 2000, the mean age of first graders rises from 10.0 years among children less than 1 km. from school to 11.4 years among children 5–6 km. from school, as might be expected, since longer distances are more daunting for younger children. But even among children less than 1 km. from school, the age at entry is much higher in rural areas (10.0 years) than in urban areas (8.3 years).

Multivariate analysis of the WMS/HICES 2000 data described in Appendix A confirms that distance from school plays an important role in determining school registration rates, that distance is far from the sole determinant of registration rates, and in particular that differences in distance to school offer only a partial explanation for differences in school attendance rates between rural and urban areas. Probit estimates employing the WMS/HICES data indicate that each additional km. of distance from the nearest primary school reduces enrollment rates by 2 to 3 percentage points, up to distances of 12 to 15 kilometers. Reducing the distance to school from 10 to 0 could thus increase regis-

Table 5.2. Among Children Aged 7 to 14 Years Old, Percent Who are Currently Registered for School, By Distance to Nearest Primary School

Distance to to Nearest Primary School in Km.	Rural		Urban		All	
	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered	Number of Obs.	Percent Registered
WMS 1995						
0	1595	22.9	3246	77.4	4841	42.3
1–2	2022	17.0	1781	68.3	3803	25.7
3–4	1773	16.0	144	80.6	1917	16.6
5–6	1354	12.0	11	56.3	1365	12.1
7–12	1178	8.8	14	68.9	1192	9.0
Over 12	356	3.8	3	19.6	359	3.8
All	**8279**	**15.4**	**5199**	**74.6**	**13478**	**23.6**
WMS 1998						
0	7574	36.9	5865	81.2	13439	50.5
1–2	8704	31.2	3883	80.0	12587	40.1
3–4	8529	24.3	568	77.5	9097	26.1
5–6	6598	19.5	78	76.7	6676	19.8
7–12	4901	11.9	3	100.0	4904	11.9
Over 12	2306	5.7	17	84.7	2323	6.1
All	38676	25.4	10450	80.5	49126	32.6
WMS 2000						
0	3680	43.6	4239	85.5	7919	57.0
1–2	5111	38.8	3827	83.1	8938	47.4
3–4	5070	32.6	454	78.9	5524	34.1
5–6	3206	24.5	61	75.7	3267	24.9
7–12	2125	16.4	9	24.3	2134	16.4
Over 12	681	8.1	8	87.6	689	8.2
All	19921	32.6	8633	83.8	28554	39.3

tration rates by 20–30 percentage points, though reducing distances from 25 to 15 km. would make little difference. The significant effects of distance to school appear primarily in the rural sample. The estimated distance effects are not diminished much by including controls for distances to other services and other community characteristics, nor the inclusion of household resource measures. They are also robust to the inclusion of EA fixed effects (verifying that the distance measures are not just picking up more general community characteristics) and to the use of instrumental variables procedures for handling potential measurement errors in the distance measures (using median distances within EA s as instruments).

As one might guess, distance plays a stronger role in inhibiting enrollments among younger children, for whom travel over long distances is more daunting and over whom parents are more anxious to keep close watch. This suggests that improving the distribution of schooling may improve schooling attainment not only by increasing the percentage of children who ever enter school, but also by facilitating an earlier age of entry, which may lead to higher ultimate attainment for reasons mentioned above.

Further insight about supply constraints is gained from the analysis of the LFS/EMIS data in Appendix B. The average number of schools per 1000 population in a woreda is approximately 0.2. Children of primary school age constitute about one fifth of the population; thus this represents about 0.2 schools per 200 children, or 1 school per 1000 children. According to the EMIS 1992, the median urban school has about 1000 students, the median rural school has around 360 students, and about 17 percent of schools are labeled as urban in the school census. Thus the simple statistics suggest that schools are still scarce, especially in rural areas.

Just because schools are scarce does not mean that building more schools will necessarily increase school registration rates, because other constraints (such as lack of income) may prevent households from taking advantage of new schools. It would thus be useful to obtain good estimates of the effect of increasing school supply on household schooling choices. This was attempted in Appendix B, which described why many caveats must be placed on the results. Taken at face value, the coefficient on the number of primary schools (offering at least grades 1–4) per 1000 population in the multivariate results suggests that doubling the number of schools offering grades 1–4 would increase enrollment rates by about 6 percent. This is a sufficiently small effect as to be implausible. Measurement error problems (associated with measuring this supply feature only at the woreda level and not at the more disaggregated community level) may help explain this, as discussed in Appendix B. It should also be noted that part of the school supply effect may be captured by the woreda population density variables, which have important effects in the rural sample. Holding the numbers of schools per 1000 population constant, children tend to live closer to the nearest primary school in woredas with higher population density. Thus the positive coefficient on population density picks up the effect of living closer to school, as well as other effects.

Another aspect of the regression results suggests the potential importance of school supply for schooling outcomes. For example, when taken together, all the variables describing school supply (which describe numbers of schools and teachers in several grade ranges, as well as some indicators of school infrastructure and teacher qualifications) seem to explain a large fraction of differences in enrollment rates across administrative regions that aren't explained by differences in the household-level variables. As described in the appendix, the aggregation of the school census data to the woreda level makes it impossible to assess the extent to which differences in school supply help explain the large rural–urban differences in registration rates. One suspects that if the supply characteristics could be measured at the community level, they would help explain a substantial fraction of rural–urban differences as well.

The analysis of the LFS/EMIS data also gives some hints that the nature and quality of the school supply—and not just the number and dispersion of schools—matters for school attendance. For example, the results seem to suggest that where secondary schools and teachers are more available, primary school enrollments are higher. (This is consistent with conjectures that parents see a higher return to primary schooling where it is easier to continue on and complete secondary school, though the results are subject to so many caveats that such a strong interpretation is not justified.) Crude indicators of the nature of school infrastructure (such as the share of primary schools that have libraries) also may matter, especially in urban areas. Given the collinear and crude nature of the data, the results should not be taken as definitive evidence that, for example, the existence of libraries per se is important for encouraging enrollment. Rather, such results indicate that in woredas that have made more investments in a variety of features of their education system, enrollment rates are higher. Further research on school system features that matter most to parents would be useful.

The Role of Income. If low incomes constitute a barrier to school attendance, then building and equipping schools will not be enough to get all children into school, and policymakers may need to allocate some scarce resources to scholarship programs. It is thus important to examine the role of household resources in determining child schooling outcomes. Ethiopian incomes are certainly low enough that they might constitute a barrier. According to the WMS/HICES 2000 data, the average child aged 7 to 14 lives in a household in which per adult equivalent consumption expenditure is the equivalent of about U.S. $176. Per adult equivalent consumption expenditure ranges from an average of U.S. $90 among children in households in the lowest quintile of the distribution, to an average of only U.S. $386 among children in households in the highest quintile of the distribution.

The simple descriptive statistics presented above (Table 3.3) suggest quite modest effects of income on school registration rates. Multivariate analysis of the WMS/HICES 1999/2000 data confirm the existence of modest household resource effects on schooling rates. In contrast to the simple statistics, the multivariate results indicate that the resource effects are smaller in urban areas than in rural areas. (The difference between the simple tables and the multivariate approach is that the multivariate approach does a better job of separating out true income effects from the effects of other household and community factors that tend to be correlated with income.) Even in rural areas, however, moving children from the bottom quintile to the top quintile of the distribution of per adult equivalent consumption expenditure increases school attendance rates by only 6–10 percentage points. Because the current expenditure variable is a noisy measure of household resources over the longer run of relevance in the variable environment of rural Ethiopia, this may understate the true effect of household resources. The results are, however, robust to the inclusion of enumeration area fixed effects, indicating that they are not just picking up the effects of differences across communities that differ in typical level of prosperity.

Two additional sets of results suggest that the importance of income constraints in producing differentiation in schooling outcomes across income classes will rise over time, at least at first, as supply constraints are relaxed. First, estimated effects of income quintile on attendance rates are much larger for younger children (ages 7 and 8) than for older children. As discussed in the appendix, this could mean either that income effects are truly larger for the younger children, who entered school age in a period of expanded school supply, or that the true income effect for all children is higher than the more modest estimates, but that measurement error biases the estimates downward more for older children. Either way, with coefficients indicating that a shift from the bottom to the top quintile would increase school attendance rates by over 18 percentage points, such results suggest that income may be becoming a more important determinant of schooling outcomes than it seems at first glance.

A second observation that suggests the role of income is growing is that the estimated effect of income quintile on school attendance rates is larger among children who live within 2 km. of the nearest primary school than among children who live further away. Among those who live further away, there is little potential for differentiation, because school availability is poor. Among those for whom schools are available, income plays a more important role in determining which children take advantage of the schools. Even among those within 2 km. of the nearest primary school, estimated income effects are modest (but may suffer from the same measurement error problems that cause overall income effects to look modest, even while the effects on the youngest children are large). Among those close to schools, moving a child from the bottom to the top quintile increases registration rates by about 10 percentage points. The biggest impact is associated with moving from the bottom to the next quintile, which increases attendance rates by close to 8 percentage points.

The Role of Opportunity Costs. Household resource levels matter for schooling choices in great part because schooling has up-front costs and households can't borrow against future education benefits to finance those costs. A potentially important component of those up-front costs is the opportunity cost of the child's time. A first question regarding opportunity costs concerns the nature of the activities that compete with schooling, thereby rendering it costly to spend time in school. Identifying the competing activities will sometimes generate ideas about how to make schooling less costly without having to create scholarship programs. For example, if the opportunity cost of a child's time is associated with work in jobs that must be done during school hours, or that must be done for long hours during certain weeks or months of the year, then changes in school schedules or curriculum may be able to reduce the opportunity cost of schooling by rendering schooling more compatible with work. Allowing children to study at night, or employing self-paced curricula that allow students to "stop out" for weeks at a time and then pick up where they left off, may improve the potential for attending school without sacrificing as much income.

A first step in assessing the likely nature of opportunity costs is to examine statistics describing the types of activities in which children are engaged. Appendix Table B.5 indicates that over half the rural boys and nearly a third of rural girls report having worked in a range of income generating activities. What the table does not detail is that almost all of these jobs involve work on the family farm. The share of children reporting income generating activities is much smaller in urban areas; but even there, a large fraction of those reporting such work report working in a household agricultural enterprise. Most of the rest are working in a household nonagricultural enterprise or self-employment. Thus few children are working in jobs that seem likely to have rigid daily hours requirements. On the other hand, many children—especially in rural areas—are engaging in work that probably has high episodic demands that might conflict with schooling for days or weeks at various points during the year. Large fractions (ranging from one fifth for rural boys to over half of urban girls) report being engaged in unpaid domestic service. No more detailed definition of this category is available, but it might include care for younger siblings and other kinds of housework. Such work is more prevalent for girls than for boys, and in urban areas.

Just because many children engage in an activity does not mean that it competes with schooling and imparts an opportunity cost to school attendance. If the work may be completed outside of school hours, the activity may not give rise to large opportunity costs. As demonstrated in Table B.5, rural children who attend school are almost as likely to report working (on the family farm) as those who are not. In urban areas, in contrast, children in school are quite a bit less likely to work a job. Interestingly, for all four groups (rural and urban boys and girls) children in school are more likely to report involvement in unpaid domestic work than children not in school. Thus the descriptive statistics paint of picture of involvement by many rural children in income-generating activities that are quite compatible with attending school. They raise somewhat more concern about the possibility that work and school are competing uses of time by urban children. Many children are also involved in a broad, ill-defined category of "unpaid domestic service," which includes many activities that seem to be compatible with schooling. Such statistics leave many questions unanswered, however, because there may be subsets of children engaged in narrower ranges of activities for which the opportunity cost of schooling is high indeed. We can hope to press further into these issues using multivariate analysis.

Several features of the multivariate results suggest that income-generating activities do compete with school for some children's time. According to results from the LFS data (as reported in Appendix B), school attendance is higher in communities in which a higher fraction of household heads is without employment (an indicator of slackness of local labor markets). This effect is strongest in urban areas. Within urban areas the estimated effect is modest, indicating that a 10 percentage point increase in the unemployment rate among household heads would be associated with a 0.6–0.8 percentage point increase in school attendance, while being associated with a reduction in child participation in work of perhaps 1.8 percentage points. This suggests that the average effect on enrollment (across all children) of increases in labor market slackness is not large, but allows for the possibility that for some subsets of children the effects are large.

Analysis of all three datasets suggests that in rural areas a better indicator of the demand for a child's time (than local unemployment rates) may be the share of household members who are adult males. In households better endowed with adult male labor, schooling rates of boys are lower, perhaps because the productivity of their time on the farm is enhanced by the presence of adult males who can direct their work and teach them. Thus, even though much work is compatible with attending school, some demands for children's time in income-generating activities probably do pull children out of school.

Several features of the multivariate results also suggest that pressing needs to help at home also keep some children out of school. Especially in urban areas, and especially for girls, those in households in which a higher fraction of members are young children are less likely to attend school. For urban girls, estimates suggest that increasing the share of young children among household members by 20 percentage points (roughly the effect of shifting one more family member into this age bracket) would reduce schooling rates by about 4 percentage points, while increasing participation in unpaid domestic work by a comparable amount.

The final observation about opportunity costs is that they seem likely to rise as development proceeds, potentially implying a growing role for income in determining schooling outcomes. As labor markets tighten and unemployment rates fall, the opportunity cost of children's time in the labor market is bound to rise. The observation that rural children in households with more adult males are less likely to attend school suggests that at least some sorts of improvements in farm productivity will increase the pull of farm work on children's time. (Fortunately, improved farm productivity may be associated with two other effects that would tend to counteract the effect of increasing opportunity costs, leaving the net effect difficult to predict. Increased productivity would also increase total household income. Furthermore, if the productivity advance is associated with new technologies that only educated farmers can take advantage of, then the perceived value of education may increase).

The Perceived Value of the Schooling on Offer. In a country where schooling rates are as low as in Ethiopia, a plausible barrier to the schooling of children is simple lack of belief in the potential benefits of schooling arising out of mere lack of exposure to education. Across all rural enumeration areas in the LFS sample, the mean and median share of household heads in the sample who are literate is around 24 percent, but this fraction in zero in almost five percent of rural enumeration areas, and ranges up to a maximum of 91 percent. Thus exposure to schooling varies greatly. This raises the possibility that for some children the most cost-effective policy that would get them into school would take the form of an education promotion campaign, or perhaps an adult literacy program, which helps their parents understand the potential importance of schooling.

Descriptive information from the WMS (Table A.9) provides some reason for concern. Among rural households reporting that they did not make use of the nearest primary school, who were asked why they don't use the nearest primary school, higher percentages responded "no experience" or "have no need of it" than "too expensive," or "too far," or "poor quality." (Half of the responses were coded as "other", however, rendering the information less than conclusive.)

Several features of the multivariate results reinforce the concern that inadequate exposure to the potential benefits of education keeps some parents from taking an interest in educating their children. One of the consistently strongest effects picked up in the estimates in Appendices A and B is the coefficient on the percentage of household heads in the child's enumeration area (not counting the child's own household head) who are literate. Coefficient estimates on the order of 0.20 in the WMS data indicate that increases of 50 percentage points in the local adult literacy rate would be associated with increases of 10 percentage points in school attendance rates, even after controlling for household resources levels, distance from school and whether or not the child's own household head is literate. Estimated effects are higher among households living closer to primary schools and households in lower income brackets. (Coefficient estimates are much higher in the LFS data, probably because the controls for household resources and community-level variation in school supply are weaker, causing the adult literacy measure to pick up some of those effects.)

Inclusion of enumeration area household head literacy rates sheds new light on the meaning of an empirical result found in datasets from many places. It is common to find that, even after controlling for income and other circumstances, children whose parents have more education are more likely to go to school. This apparent effect is often interpreted as evidence that education modifies peoples' preferences toward schooling. We see this in the current datasets in the form of large and significant coefficients on indicators of whether the child's household head (or, in the case of the DHS, his father and mother) is literate. The size of this coefficient is diminished greatly, however, when cluster average literacy rates are included in the regression, indicating that much of the association between household head or parental literacy arises not because the household head per se is literate, but because literate household heads are more likely to live in communities where many people are literate. Estimated effects of parental literacy fall even further in enumeration area fixed effects specifications, indicating that they are picking up the effects of community characteristics that are not fully captured by adult literacy rates. This says that exposure to, and beliefs about the importance of, education may be formed in a community context. It may be useful to think about exposing communities to the potential benefits of education, rather than about exposing each school aged child's parents more specifically.

Further circumstantial evidence regarding the importance of attitudes and beliefs in shaping schooling outcomes is found in the DHS dataset. Children of mothers who listen to the radio have school attendance rates at least 5 percentage points higher, even when holding constant an array of household assets, parental education, demographic characteristics of the household, and attitudes about the conditions under which is it justified for a husband to beat his wife. While the causal interpretation is unclear, it points to a potentially important association between exposure to modernity and ideas and the schooling of children.

Language. If language constitutes a barrier to schooling, then policy makers may want to continue working with reforms regarding the language of instruction. Simple tabulations such as those presented in Table 5 above indicate substantial differences in school registration rates across language groups. Because language is likely to be associated with geographic location, economic prosperity and other factors that also influence schooling rates, however, such differences cannot be attributed to language per se. Multivariate analysis employing the DHS dataset confirms that significant differences in schooling rates remain across children whose native languages differ, even after controlling for the full range of regional and household determinants available in that dataset.

The multivariate results indicate that the magnitude of the enrollment rate differences across language groups is reduced only a little by the introduction of the full set of controls. In the rural sub-sample the effect of speaking Amarigna remains large and precisely estimated even after the introduction of enumeration area fixed effects, indicating that the differences are observed even within small communities. This adds weight to the interpretation that it is language per se, and not just associated differences in culture and prosperity that cause school registration rates to differ. In urban areas, by contrast, large differences in registration rates across language groups are greatly diminished by the introduction of controls. Native language per se seems to be less a barrier to school enrollment in urban areas, where it is probably more common to speak Amharic in addition to one's native language.

Gender. Much of what the data have to say about gender differences is visible in the simple cross-tabulations of school registration rates by gender in Table 3.5. Multivariate results confirm that differences in schooling rates between boys are girls are substantial, and this is primarily a rural phenomenon. The magnitude of gender differences varies across regions, suggesting the likely importance of addressing gender issues on a region-by-region basis. Analysis of opportunity costs suggests urban girls are more likely to be drawn out of school into unpaid domestic service by the presence of younger siblings, indicating that the schooling of girls and boys would be affected differently by the provision of new child care or pre-school options in urban areas. Perhaps the most interesting observation is that the estimated effects of both school supply and population density are much larger for rural boys than for rural girls, hinting that gender differences are likely to widen before they fall as school supply expands and economic development proceeds.

Orphans. Much of what the data have to say about orphans is also evident in the simple cross-tabulations (Table 3.6). If anything, the multivariate results render the schooling gaps between children who have and have not lost parents wider and more disturbing. The results of Appendix B indicate that, after controlling for a variety of household, community and regional characteristics, children who have lost one parent are 5–6 percentage points less likely to attend school than children who have not, while the effect of having lost both parents is twice as great. The most disturbing result is that the estimated effects are as big or even larger in regressions controlling for household fixed effects. This indicates that even within households containing both natural children and children who have lost parents and now reside with adults other than their parents, children who have lost parents are less likely to attend school. This rules out the possibility that their schooling rates are lower simply because their addition to a household drives down the level of resources relative to the number of family members. In the analysis of time use, children who have lost parents are less likely to attend school, but there is no strong evidence that they are more likely to work or perform housework than other children.

Summary. Despite tremendous success in extending 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. Establishing schools in additional communities, reducing the typical distance children must travel to school, will continue to play an important role in raising enrollment rates in the coming years. Reduced distances are also likely to reduce the typical age at which children start school. Relaxing supply constraints alone will not, however, draw all children into school. Improvements in the quality of schooling offered may be important for convincing some parents to send their children to school, though the exact nature of the required improvements and likely sizes of impacts cannot be identified from the current analysis. Estimates suggest that a broad program of income transfers is unlikely to eliminate the remaining barriers to school enrollment, though more tightly targeted income transfers may be useful in some communities or groups. The need to work on the family farm, or to care for siblings and do housework, does seem to keep some children out of school. It is thus important to consider modifications to school schedules that would reduce the conflict between attending school and fulfilling responsibilities at home. A variety of evidence suggests a strong role for parental exposure to, and beliefs about, the benefits of education in shaping their decisions regarding the education of their children. Efforts to engage parents and convince them of the benefits of education thus seem likely to be important in efforts to attain the objective of Education for All.

CHAPTER 6

Suggestions for Future Data Collection

This section collates some practical suggestions for future household surveys that arise out of difficulties encountered in the preceding analysis. Because the merged WMS/HICES dataset turned out to offer the most insights relevant to schooling policy, I frame them as suggestions regarding modifications to the WMS questionnaire.

- 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 who have ever attended school, as well as more complete identification of highest grade attained.

- 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.

- Using qualitative research, extend and improve the list of categorical responses to the question about why households do not use the nearest primary school. Half the responses are currently listed as "other." 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?"

- Using qualitative research, design a question about parents' beliefs regarding the value of schooling, perhaps distinguishing the perceived value for boys and girls.

- For children who have attended school, ask the age at which they first attended primary school. Using qualitative research, design a question for children who first attended school after age 7 why they did not start school at age 7.

- Create a way of linking the individual- and household-level data to school and population census data at the level of the enumeration area or some closely related small geographic area.

- An alternative to the previous suggestion might be to create a community questionnaire to be administered in each enumeration area included in the WMS. This community questionnaire could gather key descriptive information about the schools available to community residents, as well as other community characteristics. This may be most meaningful and worthwhile in rural enumeration areas. (For guidance on the development of community questionnaires, see Frankenberg, 2000.)

- Even more might be learned about the effect of changes in school supply on schooling outcomes if a future wave of the WMS were administered in the same enumeration areas as a previous wave (though not necessarily to the same households). Such a survey design would allow observation of how schooling outcomes are changing within specific communities. Changes in schooling outcomes aggregated to the enumeration area level could be linked to local changes in school supply (assuming community-level information on historical and current school supply is gathered as well).

APPENDIX A

Analysis of the Merged WMS/HICES 2000 Data

This appendix describes the application of the empirical framework described in the main text to the analysis of the determinants of child schooling outcomes using the sub-sample of the WMS 2000 dataset that can be matched with the HICES 2000. The matched sample pertains to 16,668 households throughout Ethiopia. Matching with the HICES allows inclusion of a carefully elicited measure of household consumption expenditure, which serves as a measure of household resources. The strength of this dataset is that it allows simultaneous examination of a key feature of school supply—distance from school—and a potential determinant of demand that is of special interest to policymakers—household resources. My emphasis is on careful analysis of these two effects, as well as on trying to understand the tremendous overall differences in schooling outcomes between rural and urban areas.

The first two columns of Table A.1 report descriptive statistics, within rural and urban sub-samples, for the main variables employed in the analysis. The construction of the dependent variables is described in the main text.

Region Indicators. As discussed in the main text, the region variables are included in the multivariate analysis to pick up the effects of any systematic differences in schooling determinants across administrative regions other than differences in measures included in the regression. Note that the bulk of rural observations are from Amhara (the "left out" category), Oromiya and SNNP, while the bulk of urban observations are from Amhara, Oromiya and Addis Ababa.

Distance to Schools. The only information available in the dataset that allows some insight into the role of school supply policies in shaping child schooling outcomes are household responses to questions about the distance in kilometers to the nearest primary school. If distance from school truly has a large impact on households' willingness to school their children, then continued efforts to expand supply (in the sense of building, staffing and equipping more schools in more geographically dispersed locations) can be expected to have large benefits in expanded school attendance and increased school attainment. It is possible, however, that distance from school is not the real constraint. Building schools might not have a big impact if households perceive little benefit to education or if they are so poor that they cannot afford to buy books or to release their children from work on the family farm. It is thus valuable to derive good estimates of the impact of distance from school on schooling outcomes. The first step in this direction is to include as a potential determinant of schooling outcomes a variable measuring the distance in kilometers from the child's household to the nearest primary school.

Merely including this variable in the regression analysis does not guarantee the derivation of unbiased estimates of the intrinsic effect of distance on schooling outcomes. Econometric problems associated with the derivation of good school distance effects are discussed below.

Other Distance Variables. The household's distance from social services and economic activity may also influence the schooling of children. Aiming to control for these effects, I include measures of the distances in kilometers to the nearest post office and the nearest all-weather road.

EA-Average Variables. A household's interest in schooling children may depend on the social and economic circumstances of the community in which it resides. For example, schooling rates may depend not only on the physical distance to school, but also on what we could call the social distance to school, which is greater in communities in which there is little exposure to schooling. In some communities in Ethiopia, very few adults have ever been to school and very few households currently send children to school. In such an environment, one can imagine that there might be important spillover effects from the schooling of one household to another, as schooling in some households generates exposure to education and its benefits for that household's neighbors. An attempt to capture such spillover effects involves including as controls "EA average variables" describing the level of literacy among neighboring household heads and the level of primary school use by neighboring households.

EA average variables are calculated from WMS sample data (including observations that do not match with HICES data). They make use of the data on all households in the same enumeration area except the household of the child to which the observation pertains. Because these averages are based on samples rather than surveys, they contain sampling error.

Household Resource Variables. One of the most important strengths of the WMS/HICES data is the ability to include measures of something approximating "income." In particular, the HICES allows calculation of the household consumption expenditure for each household. For comparability across households, this measure is deflated by a regional price index (calculated by the CSA and reported in WMU, 2002). and divided by the number of "adult equivalents" in the household (using the Eastern Africa adult equivalence scale). To allow real per adult equivalent consumption expenditure to enter the determination of schooling outcomes somewhat flexibly, while keeping tables easy to read, indicators of whether the household is in the first, second, third, fourth or fifth quintile of the country-wide distribution of this measure are constructed. Indicators for quintiles 2 through 5 are included in the regressions. Quintile 1 is the "left out" category.

Household consumption expenditure variables are the preferred measures of household annual income levels, because recall problems require survey designers to use reference periods of a week or at most a month when eliciting information about income or expenditure. Consumption expenditure in the last month is likely to be a better reflection of annual income than income in the last month, in regions where incomes fluctuate with the seasons and households have ways of "smoothing" consumption relative to those income fluctuations.

Despite these reasons for preferring consumption expenditure over income measures of household resources, consumption expenditure measures are far from perfect. It is important to consider possible econometric problems associated with estimating household resource effects. We return to this concern below.

It is sometimes argued that households experience economies of scale in consumption, in the sense that the same amount of real per adult equivalent consumption expenditure goes farther when household size is larger. To allow for this possibility that the same amount of per adult equivalent consumption expenditure might finance more schooling in larger households, the adult equivalent household size is also included.

Household Head Variables. If household resource levels are well captured by the expenditure variables, then characteristics of the household head should enter primarily through their association with differences across decision makers in preferences toward schooling.

Household Structure Variables. Controlling for total household resources and needs, household composition (across adult males, adult females, school-aged children and younger children) may influence schooling choices by influencing the opportunity cost of the child's time in school. We might expect that the more young dependents a household has, the more likely a school-aged child would have to spend time caring for younger siblings, and perhaps the less likely to he or she is to attend school. Holding the number of young dependents constants, however, a child might be more likely to attend school if there are more adult women in the household who could undertake that care. Results below raise the possibility that a higher share of adult males within the household increases the opportunity cost of the child's time in the family farm or non-farm enterprise.

Child Variables. The indicator of whether the child is male is included to pick up differences in treatment of boys and girls with regard to schooling investments. Age indicators are included to allow flexibly for differences in rates of school attendance by age. In particular, given the tendency for many children to start school late, we expect to see positive and rising coefficients on these indicators, indicating that as children age (starting at age 7) they become more likely to enter school.

Econometric Difficulties in the Estimation of School Distance Effects. To derive the best possible estimates of school distance effects, we must consider four econometric concerns. First, the distance variables may pick up the effect not of distance from school (and associated pecuniary and psychological costs of having to send children further away from home), but the effect of omitted variables that are correlated with distance from school. Households that are far from the nearest school are probably isolated in other respects, and this isolation might continue to affect schooling choices even if a school were built close by. For example, the household may be isolated from markets and social services, and as a result having low expectations regarding the value of education for children. For the purposes of analyzing the potential impact on schooling outcomes of building more schools (while leaving other forms of isolation constant), we would like to purge the school distance variables of this bias. Fortunately, the dataset allows me to include the additional distance measures mentioned above. It also turns out that these measures are not highly correlated with distance to the nearest primary school, thus it appears possible to estimate their separate effects.

Second, there is reason to suspect that households' reports of distance to the nearest primary school contain substantial measurement error. Distance reports vary so much even within enumeration areas, that the authors of WMU (2002) conclude that measurement errors must be substantial. It is possible that people fail to answer questions about distance in kilometers accurately, because they are not used to thinking in these terms and have never actually measured the distances. If the measurement errors are independent and identically distributed across households, then, the measurement error tends to bias estimated coefficients downward in absolute value. That is, this measurement error would tend to cause regression estimates to pick up too little impact of distance on schooling outcomes.

Instrumental variables estimation procedures offer a potential remedy for measurement error biases. The key is to find one or more "instruments" for the erroneously measured household level distance variable that satisfy three conditions: (1) they are correlated significantly with the correctly measured household distance variables that we would like to include in the regression; (2) they are uncorrelated with the measurement error in our erroneously measured variable; and (3) they are not correlated with the errors in the schooling outcome regressions of ultimate interest. This last requirement can be restated as follows: after accounting for all the factors explicitly controlled for in the regression, the instruments should have no significant effect on the child schooling outcomes other than through the household distance to school effect that we are trying to estimate.

It can be argued that the variable "median distance to school within enumeration area sub-sample" satisfies these requirements, provided care is taken to control explicitly for various community level factors that may influence schooling outcomes. That the median and individual measures are correlated is confirmed strongly in the data, though the individual measures vary greatly around the median measures within clusters. Thus the first requirement is satisfied. If the measurement errors plaguing the household level measures are independently and identically distributed across households within the community, then the means and medians of the household errors within communities should be near zero, causing the cluster medians to be largely uncorrelated with the measurement errors. Thus under reasonable conditions the second requirement is satisfied.

The validity of EA median distances as instruments for household-level distances thus rests on the validity of the third condition, that (after controlling for all the direct controls included in the regression) the median distances have no effect on child schooling outcomes other than through the channel of individual household school attendance costs. The main reason to question this last condition would derive from the possibility that community-level distance variables might pick up important features of the culture and economy of the community, which may affect child schooling outcomes through channels unrelated to the household's cost and difficulty of getting to school. Confidence in the last condition may be increased by including good community-level controls for these community effects. The best I can do along these lines in the current dataset is to include the EA average variables shown in Table A.1, as well as EA median distances to the nearest post office and the nearest all-weather road. (Because the instruments vary only at the EA level, I cannot use EA fixed effects to control more completely for community influences.)

Third, even if current distance to school were measured accurately, it would constitute an imperfect measure of the distances from school that the household experienced over the relevant past. The five years before this survey was collected were years of much school building, especially in areas most remote from schools before. This suggests that current variation in distance from school may understate the typical differences experienced by households in access to school over the relevant past, and that coefficient estimates on school distance would tend to overstate the permanent effects of distance on schooling outcomes. There is no obvious solution to this problem. A possible way to assess the importance of this would be to identify a region of the country in which little school building took place in the recent past, and estimate the school distance effects in unconditional regressions within that sub-sample. Such efforts are beyond the scope of the current analysis.

Fourth, the effect of distance to school (and of other distance variables) on registration rates may be more complex than that implied by including the linear distance in kilometers term on the right hand side of a linear or probit regression. It is possible that the effect of increasing distance from 1 km. to 2 kms. on the probability of attending school is different from the effect of increasing distance from 9 kms. to 10 kms. I examined a number of regressions allowing for more flexible functional forms, and examined the effects of eliminating a small percentage of especially large values. I found that the effects of each of the distance variables are reasonably well approximated by splines with negative coefficients up until 12 or 14 kms., and then zero coefficients thereafter. This can be incorporated into the regressions by replacing the three distance variables by truncated versions of the same three variables, which are equal to the original variables for values less than or equal to 15, but are set to 15 for any observations in which the original variables take values above 15.

Econometric Difficulties in the Estimation of Household Resource Effects. To avoid drawing misleading inferences regarding household resource effects, we must consider five sources of potential bias. First, current consumption expenditure may be measured with error. If the measurement error is classical, it would tend to bias toward zero the estimated resource effects. It is hoped that by focusing on quintiles rather than continuous measures, this problem is reduced.

Second, imperfect regional price deflators may introduce error into the construction of real consumption expenditure measures. If the errors are independent across regions, this would introduce an attenuation bias in simple cross section regressions. As long as all households within an enumeration area face the same price levels, however, EA fixed effect estimates should be free of this bias.

Third, given fluctuations in income from year to year, current household resource levels are an imperfect measure of the lifetime household resource levels of relevance to unconditional schooling outcomes. If year-to-year income changes are positively autocorrelated, then the variance of this "measurement error" is likely to get larger, the greater is the length of time between the period in which household resources had the biggest impact on the schooling outcome and the current period. If the periods of greatest relevance were, say, when the child was 7 and 8, this suggests that measurement errors should be more severe for older children. If measurement errors are independent and identically distributed across households at any one point in time, then they should bias coefficients toward zero, and the size of this bias should be greater for older children.

It may be possible to assess the likely importance off this bias in regressions of whether a child ever attended school. For the reasons just described, measurement error biases toward zero should be higher (and estimated income effects lower) for older children (at least above some age at which most parents contemplate sending children to school). If we see income effects hold steady across ages, then either measurement error problems are unimportant or they are compensated by a pattern of true income effects that were greater for children born earlier. The latter possibility seems slim. The recent past is a period of much building of schools. As schools become more accessible, it seems likely that income effects would become more important rather than less. When there are no schools around, income doesn't matter. Only when schools are available does income matter. Thus if income effects hold steady across ages, measurement error is not likely to be a big problem in the estimation of income effects on ever attendance. This would mean that current income does a reasonably good job of picking up the effect of longer-run resource levels. If, however, current resource effects are smaller for older children, we do not know whether this is because of more severe measurement error problems or because the true effects are indeed smaller.

Fourth, household consumption expenditure may be correlated with unobserved levels of local economic development, which both boosts typical household incomes and renders households more inclined to send their children to school (even if their own incomes did not rise). If the policy experiment we have in mind is an economic development experiment, then we might not consider this a bias. If, however, our policy experiment pertains to targeted cash transfers, then we would prefer to purge our household resource estimates of this effect. Introducing community fixed effects is useful for this purpose. In the community fixed effects regressions, the household resource effect is estimated off of variation in household resources within communities, and thus holds the level of community development constant.

Finally, current household resources may be correlated with unobserved household characteristics that influence both labor supply and income, on the one hand, and the schooling of children, on the other. For example, it may be correlated with unobserved attitudes and expectations regarding the potential for individuals and households to get ahead in life if they work hard. Higher levels of such hope would tend both to increase income and increase the probability of a child continuing in school. It might thus tend to introduce an upward bias in the estimation of household resource effects. Unfortunately, there is little to be done to eliminate this potential bias. (Some authors attempt to use an instrumental variables procedure to get around this, but the methods rest on the validity of assumptions that are difficult to justify. For example, they may rest very strongly on the assumption that the preference-related effects of parental age are linear, while the effects of parental age on household resources are quadratic.)

Basic Multivariate Results. Columns 3 through 7 of Table A.1 report estimates of coefficients in regressions relating the probability that a child is registered for school to some or all of the potential determinants listed in Table 1. All specifications in Table A.1 employ the entire sample (both rural and urban children). Columns 3 through 5 relate to probit estimation. Comparison of the third and fourth columns demonstrates how much of the overall difference between urban and rural areas may be explained by the additional controls included in the second regression. Comparison of columns 4 and 5 demonstrates the role played by the "community variables" derived from within-sample cluster averages. The 6th column demonstrates what happens to the estimates of this basic model when effects are estimated by OLS rather than by probit methods, while the final column demonstrates what happens when enumeration area fixed effects are introduced into the OLS specification. (The motivation for this is discussed in the main text and above.)

The main observations from Table A.1 are the following:

- While the inclusion of additional controls reduces the estimated rural–urban difference, the difference remains great. Additional regressions not reported here suggest that (1) inclusion of region indicators has very little effect on the estimated rural–urban difference; and (2) the inclusion of distance measures play a larger role in reducing the size of the urban coefficient than does the inclusion of the consumption expenditure quintile indicators. Thus differences in school supply have a significant role to play in explaining differences in school registration between rural and urban areas, but by no means explain the entire difference.

- After controlling for community, household and child-level variables available in this dataset, school enrollment rates tend to be higher in the regions of Benishangul-Gumuz, Gambela and Harari, and lower in Affar, Somali, Dire Dawa and possibly SNNP than in other regions.

- The functional form assumption differences underlying the probit and OLS estimation methods play a troublingly large role in shaping estimated effects of right hand side variables on the probability of a child being registered for school. In many applications the two procedures give very similar predicted probability effects. But in this application, comparison of the third and fourth columns indicates that the shift from probit to OLS specifications makes a big difference. Of greatest concern is the observation that estimated impacts of distance and quintile variables are smaller in the OLS specification than in the probit specification. Standard arguments would suggest that the probit specification is superior.

- Probit specifications indicate a large effect of distance to school on schooling outcomes. A coefficient of –.02, for example, indicates that for each additional km. of distance between the child's home and the nearest primary school, the probability of registration falls by 2 percentage points. This suggests that for children living 5 km. from the nearest school, building a school near by could increase registration rates by 10 percentage points—a large improvement on a very low base registration rate.

- The apparent effect of distance to school is reduced by inclusion of the additional controls for community characteristics in column 3, but not by much.

- While the OLS estimate of the distance effect is smaller than the probit estimate, it is still statistically significant. More important, it is not diminished by the inclusion of community fixed effects. This suggests that the effects of distance on schooling outcomes are observed within communities (between households that live closer to and further from the same school) as well as across communities. (Note also that attenuation bias associated with measurement error on the distance measures would be aggravated by fixed effects estimation. Thus finding a significant effect, despite the stronger bias toward zero, bolsters the conclusion that distance from school matters.)

- The coefficients on the EA average variables suggest strong "neighborhood effects." That is, holding a household's individual circumstances constant, improving the adult literacy or schooling use among its neighbors tends to increase the probability of its children attending school by a substantial amount. For example, increasing the literacy rate among neighboring household heads from 0 to 100 percent would increase child schooling rates by 21 percent, even when holding constant the child's household resources and distance from school.

- Household resources (as measured by the per adult equivalent consumption expenditure quintile variables) also play a role in determining school registration, suggesting potential interest in "demand side" education policies. Increasing resource levels from the quintile 1 levels to quintile 2 levels (which is not very much, given the distribution of income in Ethiopia), increases school registration rates by around 4 percentage points. Increasing from quintile 1 levels to quintile 5 levels appears to increase registration rates by 8 or 9 percentage points. Again, the inferior OLS specification finds smaller effects. Inclusion of community fixed effects increases estimated household resource effects. Thus these variables are not simply picking up differences across communities in the level of economic development.

- As in most studies of education, children in households with literate household heads are much more likely to be registered for school than those in households with illiterate heads.

- Interestingly, however, the size of the estimated household head literacy effect drops by half when enumeration area fixed effects are introduced. This indicates that the higher household head literacy effects were in part picking up the effects of living in more literate (and perhaps more developed) communities.

- Household structure also matters. Where there are more young dependents or more adult males, children are less likely to register for school. Why having more adult males in the household reduces school attendance is not entirely clear. It may be that with more adult males in the household family enterprises become feasible, increasing the opportunity cost of the child's time. Where there are more adult females, who may substitute for school-aged children in caring for the younger dependents, children are more likely to register for school. Opportunity costs in child care thus appear important in shaping schooling choices.

- Males enjoy higher school registration rates than girls, all else equal, by 7 to 10 percentage points.

- Registration rates rise with age until about 12 years of age, reflecting late age of entry to school in Ethiopia.

The aim of Table A.2 is to examine the robustness of the "distance to school" effects to variation in the treatment of functional form and measurement error. In all columns of this table, the measures of distance employed are truncated at 15 kms. That is, whereas the distance measures used in Table A.1 contain some observations as high as 50, 70 or 90 km., all such observations are truncated to 15 km. in Table A.2. This essentially imposes the functional form assumption that while each km. of distance up to 15 km. reduces school registration rates by the same number of percentage points, additional increases in distance beyond 15 km. have no effect on registration rates. Such a functional form assumption was motivated by examination of nonparametric fits of the bivariate relationships between school registration rates and individual distance measures. As in Table A.1, Table A.2 compares probit specifications with and without the cluster average community variables, and OLS with and without community fixed effects. Table A.2 also introduces into the linear models the instrumental variables estimates, in which community median distance variables are employed as instruments for the household-level distance variables, in the hopes of reducing measurement error biases.

The main observations from Table A.2 are as follows:

- Truncation serves to increase a little the estimated impact of distance (up to 15 km.) on school registration rates (as seen by comparing columns 1 and 2 of Table A.2 to columns 4 and 5 of Table A.1).

- OLS still finds smaller distance effects than the probit specifications, but the differences are smaller.

- The use of instrumental variables as a remedy for measurement error has little impact on estimated coefficients.

- The use of community fixed effects continues to have little impact on estimated distance effects.

- Taken all together, the results support the conclusion (which is of little surprise) that school supply is still an important constraint on school attendance in Ethiopia.

Table A.3 presents estimates of school registration equations that are done separately for rural and urban sub-samples. The main observations from Table A.3 are as follows:

- While distance to school plays an important role in explaining schooling outcomes in rural areas, it does not appear to play such a role in urban areas. (It is possible that attenuation bias is stronger in the urban than rural sub-sample, because the variation in true distances is much less. Thus comparable measurement errors would lead to larger biases.)

- EA average effects also appear stronger in rural than urban areas. This makes sense. Greater isolation in rural areas would tend to imply a much larger role of a household's immediate neighbors in shaping its perceptions of the world and its actual economic and social circumstances.

- Household resource effects also appear stronger in rural than urban areas.

- Registration rates are significantly lower in urban households headed by males than in urban households headed by females, though such differences are not picked up in rural regressions (where there are fewer female-headed households) or in the entire sample.

- Household head literacy plays an important role in both urban and rural areas. The effect of household head literacy is diminished by the inclusion of enumeration areas fixed effects in both rural and urban areas.

- The presence of younger siblings is a greater deterrent to school attendance for children in urban areas than in rural areas.

- Gender differences in registration rates are much more pronounced in rural than in urban areas.

Tables A.4 and A.5 present regressions relating a set of dependent variables that offer a more "dynamic" look at children's movement into and through school to the same determinants as in the other tables, first for rural sub-samples and then for urban sub-samples. The first column examines the determinants of whether a child has ever attended school (according to the imperfect measure that may be constructed in this dataset), while the remaining columns examine the association between the determinants and rates of dropout, repeat, advance and "unclear" (as defined in the main text) among children who were registered for school last year.

As discussed in the text, great care should be exercised in comparing coefficients across columns of these tables, because the restriction of samples to children who attended school last year introduces problems of both interpretation and estimation (both of which probably suggest that we should see smaller effects in absolute value of distance and resource variables, even when their true effect on transition rates is just as important as their effects on ever attendance rates). The main observations from Tables A.4 and A.5 are:

- The probit results for whether the child has ever attended school look very similar to probit results for whether the child is currently registered for school. This is no surprise, given that the vast majority of Ethiopian children who have ever attended school are currently in school.

- The point estimates of coefficients on almost all potential determinants are much smaller in absolute value in the "transition regressions" (pertaining to the dependent variables dropout, repeat, advance and unclear) than in the regressions pertaining to whether children have ever attended school or are currently registered for school. Unfortunately, we do not know whether this is because the determinants truly play a smaller role in influencing transition probabilities than they do in shaping the probabilities of beginning school, or whether it is a statistical artifact arising out of the endogenous selection of the samples, and the conditioning on past school attendance, which causes variables to pick up more limited effects even when well estimated.

Table A.6 presents estimates of probit regressions, for whether a child has ever attended school, separately for sub-samples of children within narrower age categories. This is motivated in part by the notion that measurement error bias (associated with using current distance and household resource measures as measures of the lifetime levels that shaped current schooling outcomes) should likely be more severe for older children. The main observations from Table A.6 are:

- Distance from school appears to play a more important role in determining whether younger children have ever attended school than in determining whether older children have ever attended school. It seems reasonable that safety and fatigue concerns associated with long walks to school may be more daunting for younger children, causing household further from schools to delay longer in sending children to school. Thus this could reflect a "real" difference. Measurement error arising out of the difference between current distance to school and the distance to school that prevailed when a child was, say, age 7, may be larger for older children. Thus the smaller coefficients may simply reflect larger attenuation bias.

- Estimated household resource effects diminish as the age of the children in the sample increases. This is consistent with two possible explanations (or a combination of the two). The first has to do with measurement error. If the measurement error is—for reasons discussed above—independently and identically distributed across households and more severe for children who faced their main education decisions further in the past—then we would expect downward biases in estimated resource effects that are more severe for older children. The second potential explanation is that resources indeed play a more important role in determining the schooling of younger children than they did for older children, perhaps because school supply constraints have been relaxed.

- Gender differences are larger for older children, essentially pointing to the fact that gender differences cannot emerge until children start attending school.

Table A.7 estimates probit regressions, for whether a child is currently registered for school, separately for six sub-samples of rural children: those living 2 kms. or less from the nearest primary school, those living more than 2 kms. from the nearest primary school, those in the first and second quintiles of per adult equivalent consumption expenditure, those in the third, fourth and fifth quintiles, boys and girls. The main observations from Table A.7 are:

- Strong differences in enrollment rates across administrative regions are shared across sub-samples.

- Distance from school matters more among households closer to school than further away, as is consistent with what was said above about the functional form of the relationship between distance and registration rates.

- Literacy rates among household heads are more strongly associated with school registration among households that live within 2 km. of a school and among lower income households.

- Expenditure quintile seems to matter more among households that are closer to schools.

- Increasing household resources have a greater impact on the registration of boys than girls.

Subsidiary Observations. After asking how far the nearest primary school lies from the household, and whether the household uses this facility, it asks those households that do not use the nearest primary school to indicate why not. Table A.10 contains frequency distributions of the responses, separately for rural and urban sub-samples. The main observations from Table A.8 are:

- A very large share of responses is coded as "other," rendering inferences highly uncertain.

- In both rural and urban areas, large fractions of the responses that can be coded fall in the category of "have no need of it." This is especially striking because the sample is of children of primary school age! I have also confirmed that most households responding that they do not need the school have children who are not attending school (it is not, for example, that they don't need a public school because they are going to a better private school).

- A significant number of rural responses are also coded as "no experience," increasing concern about the communities where there is no school, few children attend school and few parents are literate. Education campaigns may be needed to begin convincing parents that education is valuable.

- Responses that the school is too far away are somewhat more prevalent in rural areas than urban, while responses that the school is too expensive or too lacking in quality are somewhat more frequent in urban areas, but none of these describe the responses of large fractions of the population.

Table A.1. Descriptive Statistics and Estimates of Probability Derivatives Employing the WMS/HICES 2000
Sample: All Children 7 to 14 Years Old
Dichotomous Dependent Variable: Whether Child is Registered for School

			Estimated Changes in Probability Associated with Change in Indicated Regressor, Various Methods[1]				
	Rural Means (Std. Dev.)	Urban Means (Std. Dev.)	Probit	Probit	Probit	OLS	EA Fixed Effects
Number of Observations	9333	8076	17,409	17,409	17,409	17,409	17,409
Urban residence[2]			0.509*	0.396*	0.305*	0.288*	
			(0.013)	(0.017)	(0.020)	(0.017)	
Region of residence (excluded category is Amhara):[2]							
Tigray	0.066	0.080		0.014	0.008	0.005	
				(0.028)	(0.027)	(0.019)	
Affar	0.003	0.008		–0.186*	–0.096*	–0.023	
				(0.042)	(0.037)	(0.026)	
Oromiya	0.394	0.298		–0.034	–0.044*	–0.023	
				(0.020)	(0.019)	(0.013)	
Somali	0.010	0.034		–0.316*	–0.270*	–0.167*	
				(0.035)	(0.032)	(0.024)	
Benshangul-Gumuz	0.011	0.007		0.123*	0.089*	0.084*	
				(0.033)	(0.032)	(0.025)	
SNNPR	0.249	0.113		–0.017	–0.049*	–0.026	
				(0.023)	(0.022)	(0.016)	
Gambela	0.002	0.004		0.208*	0.170*	0.154*	
				(0.043)	(0.045)	(0.037)	
Harari	0.001	0.008		0.115*	0.121*	0.121*	
				(0.032)	(0.025)	(0.020)	
Addis Ababa	0.001	0.227		0.049	0.092*	0.070*	
				(0.028)	(0.023)	(0.017)	
Dire Dawa	0.002	0.020		–0.184*	–0.130*	–0.078*	
				(0.039)	(0.036)	(0.026)	
Distance in km. to nearest:							
Primary school	3.467(3.459)	0.868(1.214)		–0.026*	–0.020*	–0.005*	–0.005*
				(0.005)	(0.004)	(0.001)	(0.001)
Post office	20.927(17.540)	5.128(13.044)		–0.002*	–0.002*	–0.001*	0.001*
				(0.001)	(0.001)	(0.000)	(0.001)
All-weather road	10.743(14.345)	0.415(1.693)		0.000	–0.001	0.000	0.000
				(0.001)	(0.001)	(0.001)	(0.001)
Share of households in EA:							
Using nearest school	0.482(0.254)	0.654(0.217)			0.338*	0.275*	
					(0.029)	(0.022)	
With literate head	0.245(0.167)	0.579(0.192)			0.212*	0.161*	
					(0.037)	(0.028)	
Per adult equivalent consumption expenditure quintiles (excluded category is quintile 1):[2]	0.253	0.192		0.043*	0.039*	0.025*	0.037*
				(0.014)	(0.013)	(0.010)	(0.009)
Quintile 2	0.206	0.170		0.062*	0.058*	0.038*	0.063*
				(0.015)	(0.015)	(0.011)	(0.010)
Quintile 3	0.168	0.169		0.079*	0.073*	0.046*	0.063*
				(0.016)	(0.016)	(0.012)	(0.011)
Quintile 4	0.096	0.223		0.090*	0.081*	0.052*	0.084*
				(0.017)	(0.016)	(0.011)	(0.011)
Quintile 5	5.291(1.718)	5.470(2.239)		0.010*	0.009*	0.005*	0.010*
				(0.003)	(0.003)	(0.002)	(0.002)

Table A.1 (continued)

			Estimated Changes in Probability Associated with Change in Indicated Regressor, Various Methods[1]				
	Rural Means (Std. Dev.)	Urban Means (Std. Dev.)	Probit	Probit	Probit	OLS	EA Fixed Effects
Size of household in adult equivalents							
Characteristics of household head:							
Male[2]	0.807	0.659		–0.047*	–0.028	–0.019	–0.016
				(0.021)	(0.021)	(0.013)	(0.013)
Age in years	46.290(12.278)	45.005(12.298)		0.012*	0.012*	0.008*	0.007*
				(0.002)	(0.002)	(0.002)	(0.002)
Age squared				0.000*	0.000*	0.000*	0.000*
				(0.000)	(0.000)	(0.000)	(0.000)
Literate[2]	0.230	0.604		0.138*	0.113*	0.082*	0.052*
				(0.013)	(0.013)	(0.009)	(0.008)
Number of spouses	0.794(0.440)	0.641(0.492)		0.015	0.013	0.007	0.013
				(0.019)	(0.019)	(0.012)	(0.011)
Share of household members who are:							
Under 7 years old	0.201(0.148)	0.136		–0.139*	–0.135*	–0.100*	–0.066*
				(0.047)	(0.048)	(0.033)	(0.030)
Male and over 15 years	0.217(0.124)	0.214		–0.105*	–0.105*	–0.075*	–0.098*
				(0.052)	(0.053)	(0.036)	(0.031)
Female and over 15 yrs.	0.228(0.108)	0.285		0.102*	0.100*	0.068*	–0.009
				(0.051)	(0.051)	(0.034)	(0.032)
Child characteristics (excluded categories are female and age 7):[2]							
Male	0.514	0.474		0.102*	0.103*	0.069*	0.073*
				(0.010)	(0.010)	(0.007)	(0.006)
Age 8	0.148	0.122		0.137*	0.135*	0.098*	0.088*
				(0.015)	(0.015)	(0.012)	(0.011)
Age 9	0.134	0.115		0.216*	0.217*	0.167*	0.161*
				(0.013)	(0.013)	(0.012)	(0.011)
Age 10	0.122	0.136		0.252*	0.254*	0.203*	0.196*
				(0.013)	(0.013)	(0.012)	(0.011)
Age 11	0.096	0.098		0.253*	0.253*	0.200*	0.197*
				(0.013)	(0.013)	(0.013)	(0.012)
Age 12	0.137	0.154		0.277*	0.279*	0.224*	0.222*
				(0.013)	(0.013)	(0.012)	(0.011)
Age 13	0.104	0.124		0.245*	0.247*	0.199*	0.198*
				(0.015)	(0.015)	(0.014)	(0.012)
Age 14	0.109	0.138		0.255*	0.254*	0.205*	0.200*
				(0.014)	(0.014)	(0.013)	(0.012)

[1] For continuous regressors, the estimates are of the derivative of the probability of the dichotomous dependent variable equalling one with respect to the regressor, evaluated at the means of all right hand side variables. For dichotomous regressors, the estimates are of the change in probability as the variable is changed from zero to one, while holding all other right hand side variables at their means. Robust standard errors are shown in parentheses.

[2] These regressors are dichotomous, taking the value 1 if the indicated condition is true, and zero otherwise. They capture differences in the dependent variable between the indicated category and the excluded category. Where no excluded category is explicitly mentioned in the table, it is the opposite of the category indicated.

* Asterisks identify coefficient estimates that are significantly different from zero at the two-tailed five percent level.

Table A.2. Estimates of Probability Derivatives Employing the WMS/HICES 2000
Sample: All Children 7 to 14 Years Old
Dichotomous Dependent Variable: Whether Child is Registered for School

	Estimated Changes in Probability Associated with Change in Indicated Regressor, Various Methods[1]				
	Probit	Probit	OLS	Instrumental Variables	EA Fixed Effects
Number of Observations	17,409	17,409	17,409	17,409	17,409
Urban residence[2]	0.329*	0.255*	0.225*	0.220*	
	(0.020)	(0.022)	(0.020)	(0.024)	
Region of residence (excluded category is Amhara):[2]					
Tigray	0.009	0.005	0.003	0.002	
	(0.028)	(0.026)	(0.018)	(0.018)	
Affar	–0.202*	–0.115*	–0.050*	–0.054*	
	(0.040)	(0.036)	(0.024)	(0.024)	
Oromiya	–0.039*	–0.047*	–0.025	–0.025*	
	(0.020)	(0.019)	(0.013)	(0.013)	
Somali	–0.334*	–0.287*	–0.177*	–0.178*	
	(0.034)	(0.032)	(0.026)	(0.025)	
Benshangul-Gumuz	0.107*	0.078*	0.074*	0.076*	
	(0.032)	(0.031)	(0.024)	(0.025)	
SNNPR	–0.026	–0.054*	–0.034*	–0.032*	
	(0.023)	(0.022)	(0.015)	(0.016)	
Gambela	0.190*	0.157*	0.137*	0.138*	
	(0.045)	(0.046)	(0.037)	(0.037)	
Harari	0.085*	0.099*	0.096*	0.095*	
	(0.030)	(0.024)	(0.019)	(0.020)	
Addis Ababa	0.041	0.084*	0.062*	0.060*	
	(0.027)	(0.023)	(0.016)	(0.016)	
Dire Dawa	–0.184*	–0.132*	–0.086*	–0.083*	
	(0.039)	(0.037)	(0.026)	(0.026)	
Distance in km (truncated at 15 km.) to nearest:					
Primary school	–0.033*	–0.026*	–0.015*	–0.012*	–0.015*
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Post office	–0.009*	–0.007*	–0.007*	–0.008*	0.001
	(0.002)	(0.002)	(0.001)	(0.003)	(0.003)
All-weather road	–0.003	–0.002	–0.002	–0.002	0.000
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Share of households in EA:					
Using nearest school					
With literate head		0.319*	0.241*	0.248*	
		(0.029)	(0.022)	(0.023)	
		0.202*	0.150*	0.148*	
		(0.037)	(0.027)	(0.027)	
Per adult equivalent consumption expenditure quintiles (excluded category is quintile 1):[2]					
Quintile 2	0.043*	0.040*	0.026*	0.027*	0.037*
	(0.013)	(0.013)	(0.010)	(0.010)	(0.009)
Quintile 3	0.061*	0.057*	0.038*	0.038*	0.063*
	(0.015)	(0.015)	(0.011)	(0.011)	(0.010)
Quintile 4	0.079*	0.072*	0.045*	0.046*	0.062*
	(0.016)	(0.016)	(0.012)	(0.012)	(0.011)

Table A.2 (continued)

	Estimated Changes in Probability Associated with Change in Indicated Regressor, Various Methods[1]				
	Probit	Probit	OLS	Instrumental Variables	EA Fixed Effects
Quintile 5	0.089*	0.081*	0.051*	0.051*	0.083*
	(0.017)	(0.016)	(0.011)	(0.011)	(0.011)
Size of household in adult equivalents	0.010*	0.008*	0.004*	0.004*	0.010*
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Characteristics of household head:					
Male[2]	–0.042	–0.025	–0.018	–0.018	–0.016
	(0.021)	(0.021)	(0.014)	(0.014)	(0.013)
Age in years	0.012*	0.012*	0.008*	0.009*	0.007*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age squared	0.000*	0.000*	0.000*	0.000*	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Literate[2]	0.134*	0.112*	0.079*	0.079*	0.051*
	(0.013)	(0.013)	(0.009)	(0.009)	(0.008)
Number of spouses	0.016	0.014	0.008	0.009	0.013
	(0.019)	(0.019)	(0.012)	(0.012)	(0.011)
Share of household members who are:					
Under 7 years old	–0.138*	–0.134*	–0.098*	–0.098*	–0.068*
	(0.048)	(0.048)	(0.033)	(0.033)	(0.030)
Male and over 15 years	–0.117*	–0.114*	–0.079*	–0.079*	–0.099*
	(0.052)	(0.052)	(0.035)	(0.035)	(0.031)
Female and over 15 yrs.	0.094	0.094	0.061	0.059	–0.009
	(0.051)	(0.051)	(0.034)	(0.034)	(0.032)
Child characteristics (excluded categories are female and age 7):[2]					
Male	0.103*	0.104*	0.070*	0.070*	0.073*
	(0.010)	(0.010)	(0.007)	(0.007)	(0.006)
Age 8	0.137*	0.135*	0.098*	0.099*	0.088*
	(0.015)	(0.015)	(0.012)	(0.012)	(0.011)
Age 9	0.216*	0.217*	0.168*	0.167*	0.161*
	(0.013)	(0.013)	(0.012)	(0.012)	(0.011)
Age 10	0.251*	0.253*	0.202*	0.201*	0.196*
	(0.013)	(0.013)	(0.012)	(0.012)	(0.011)
Age 11	0.250*	0.252*	0.200*	0.200*	0.196*
	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)
Age 12	0.276*	0.277*	0.223*	0.223*	0.221*
	(0.013)	(0.013)	(0.012)	(0.012)	(0.011)
Age 13	0.245*	0.246*	0.198*	0.198*	0.198*
	(0.014)	(0.015)	(0.014)	(0.014)	(0.012)
Age 14	0.255*	0.254*	0.205*	0.205*	0.200*
	(0.014)	(0.014)	(0.013)	(0.013)	(0.012)

[1,2] See Table A.1, notes 1 and 2.

* Asterisks identify coefficient estimates that are significantly different from zero at the two-tailed five-percent level.

Table A.3. Estimates of Probability Derivatives Employing the WMS/HICES 2000
Samples: Rural and Urban Children 7 to 14 Years Old
Dichotomous Dependent Variable: Whether Child is Registered for School

	Estimated Changes in Probability Associated with Change in Indicated Regressor, Various Sub-samples and Methods[1]					
	Rural			Urban		
	Probit	OLS	EA Fixed Effects	Probit	OLS	EA Fixed Effects
Number of Observations	9,333	9,333	9,333	8,076	8,076	8,076
Region of residence (excluded category is Amhara):[2]						
Tigray	–0.008	–0.015		0.011	0.010	
	(0.031)	(0.033)		(0.020)	(0.019)	
Affar	–0.122*	–0.027		–0.040	–0.038	
	(0.034)	(0.030)		(0.042)	(0.048)	
Oromiya	–0.065*	–0.054*		–0.016	–0.011	
	(0.021)	(0.020)		(0.017)	(0.015)	
Somali	–0.187*	–0.132*		–0.185*	–0.194*	
	(0.023)	(0.024)		(0.040)	(0.038)	
Benshangul-Gumuz	0.126*	0.127*		–0.012	0.004	
	(0.038)	(0.035)		(0.028)	(0.028)	
SNNPR	–0.034	–0.020		–0.077*	–0.070*	
	(0.022)	(0.021)		(0.027)	(0.021)	
Gambela	0.328*	0.308*		–0.021	–0.016	
	(0.053)	(0.039)		(0.035)	(0.031)	
Harari	0.066	0.103*		0.059*	0.057*	
	(0.035)	(0.033)		(0.015)	(0.017)	
Addis Ababa	0.083	0.055		0.024	0.020	
	(0.042)	(0.042)		(0.018)	(0.016)	
Dire Dawa	–0.110*	–0.080*		–0.110*	–0.111*	
	(0.031)	(0.032)		(0.043)	(0.040)	
Distance in km. to nearest:						
Primary school	–0.023*	–0.004*	–0.007*	0.001	0.001	0.002
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
Post office	–0.001	–0.001	0.001	–0.002*	–0.003*	0.002*
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)
All-weather road	–0.001	–0.001	0.000	–0.003	–0.006	0.000
	(0.001)	(0.000)	(0.001)	(0.003)	(0.006)	(0.002)
Share of households in EA:						
Using nearest school	0.366*	0.370*				
	(0.032)	(0.029)		0.075*	0.080*	
With literate head	0.098*	0.108*		(0.028)	(0.030)	
	(0.046)	(0.044)		0.149*	0.165*	
				(0.028)	(0.031)	
Per adult equivalent consumption expenditure quintiles (excluded category is quintile 1):[2]						
Quintile 2	0.052*	0.039*	0.049*	0.013	0.017	0.027*
	(0.015)	(0.013)	(0.014)	(0.012)	(0.014)	(0.013)
Quintile 3	0.068*	0.054*	0.075*	0.029*	0.033*	0.050*
	(0.018)	(0.015)	(0.016)	(0.013)	(0.015)	(0.013)
Quintile 4	0.081*	0.063*	0.080*	0.039*	0.041*	0.050*
	(0.020)	(0.017)	(0.017)	(0.013)	(0.015)	(0.013)
Quintile 5	0.082*	0.066*	0.107*	0.043*	0.044*	0.070*
	(0.023)	(0.020)	(0.020)	(0.012)	(0.014)	(0.013)

Table A.3 (continued)

	Estimated Changes in Probability Associated with Change in Indicated Regressor, Various Sub-samples and Methods[1]					
	Rural			Urban		
	Probit	OLS	EA Fixed Effects	Probit	OLS	EA Fixed Effects
Size of household in adult equivalents	0.007 (0.004)	0.006 (0.004)	0.009* (0.003)	0.006* (0.003)	0.005* (0.002)	0.010* (0.002)
Characteristics of household head:						
Male[2]	–0.004 (0.029)	0.002 (0.024)	–0.007 (0.022)	–0.037* (0.014)	–0.042* (0.015)	–0.030* (0.015)
Age in years	0.006* (0.003)	0.005* (0.002)	0.006* (0.002)	0.010* (0.002)	0.010* (0.002)	0.009* (0.002)
Age 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)
Literate[2]	0.076* (0.016)	0.068* (0.014)	0.049* (0.013)	0.084* (0.012)	0.088* (0.012)	0.062* (0.010)
Number of spouses	0.008 (0.025)	0.007 (0.020)	0.018 (0.018)	0.011 (0.014)	0.008 (0.014)	0.007 (0.013)
Share of household members who are:						
Under 7 years old	–0.042 (0.055)	–0.038 (0.048)	0.000 (0.045)	–0.174* (0.040)	–0.185* (0.043)	–0.180* (0.038)
Male and over 15 years	–0.077 (0.060)	–0.077 (0.054)	–0.078 (0.049)	–0.070 (0.045)	–0.065 (0.045)	–0.111* (0.039)
Female and over 15 yrs.	–0.011 (0.063)	0.008 (0.057)	–0.035 (0.055)	0.048 (0.041)	0.044 (0.041)	–0.022 (0.037)
Child characteristics (excluded categories are female and age 7):[2]						
Male	0.113* (0.012)	0.098* (0.011)	0.100* (0.009)	0.039* (0.008)	0.039* (0.008)	0.040* (0.008)
Age 8	0.127* (0.023)	0.086* (0.015)	0.071* (0.016)	0.069* (0.010)	0.111* (0.019)	0.109* (0.016)
Age 9	0.228* (0.023)	0.162* (0.015)	0.152* (0.016)	0.101* (0.008)	0.165* (0.018)	0.170* (0.015)
Age 10	0.296* (0.024)	0.219* (0.017)	0.206* (0.017)	0.107* (0.008)	0.177* (0.018)	0.180* (0.015)
Age 11	0.299* (0.026)	0.219* (0.019)	0.216* (0.018)	0.106* (0.008)	0.174* (0.018)	0.170* (0.016)
Age 12	0.331* (0.024)	0.250* (0.017)	0.246* (0.016)	0.115* (0.008)	0.186* (0.018)	0.188* (0.015)
Age 13	0.322* (0.026)	0.241* (0.020)	0.242* (0.018)	0.088* (0.009)	0.145* (0.019)	0.151* (0.015)
Age 14	0.344* (0.025)	0.259* (0.019)	0.246* (0.018)	0.087* (0.009)	0.143* (0.019)	0.147* (0.015)

[1,2] See Table A.1, notes 1 and 2.

* * Asterisks identify coefficient estimates that are significantly different from zero at the two-tailed five-percent level.

Table A.4. Estimates of Probability Derivatives Employing the WMS/HICES 2000
Samples: Rural Children 7 to 14 Years Old
Dichotomous Dependent Variables Related to Progress Through School

	Estimated Changes in Probability Associated with Change in Indicated Regressor, from Probit estimates, Various Samples and Dependent Variables[1]				
	Entire Sample	Those who registered for school last year			
	Ever attended school?	Dropped out?	Repeated?	Advanced?	Unclear?
Number of Observations	9,333	2,423	2,455	2,455	2,423
Region of residence (excluded category is Amhara):[2]					
Tigray	–0.020	–0.039*	0.002	0.046	–0.039
	(0.030)	(0.014)	(0.060)	(0.061)	(0.021)
Affar	–0.134*	–0.009	0.154	–0.190	0.035
	(0.036)	(0.034)	(0.132)	(0.105)	(0.065)
Oromiya	–0.064*	0.018	–0.043	0.009	0.030
	(0.021)	(0.019)	(0.025)	(0.034)	(0.023)
Somali	–0.209*		0.029	0.054	–0.020
	(0.022)		(0.100)	(0.102)	(0.022)
Benshangul-Gumuz	0.120*	–0.036*	0.067	–0.045	0.004
	(0.039)	(0.014)	(0.050)	(0.048)	(0.020)
SNNPR	–0.034	0.001	0.051	–0.055	–0.029
	(0.022)	(0.016)	(0.029)	(0.035)	(0.023)
Gambela	0.331*	–0.037*	0.010	0.021	–0.020
	(0.050)	(0.014)	(0.051)	(0.060)	(0.025)
Harari	0.073*	–0.007	0.005	0.015	–0.061*
	(0.034)	(0.022)	(0.045)	(0.057)	(0.023)
Addis Ababa	0.076	–0.041*	–0.066*	0.133*	0.056
	(0.043)	(0.021)	(0.031)	(0.037)	(0.045)
Dire Dawa	–0.103*	0.050	–0.002	–0.064	0.002
	(0.032)	(0.038)	(0.061)	(0.080)	(0.002)
Distance in km. to nearest:					
Primary school	–0.024*	0.002	–0.001	–0.001	0.000
	(0.003)	(0.002)	(0.003)	(0.004)	(0.000)
Post office	–0.001	0.000	0.000	0.000	0.001*
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
All-weather road	–0.001	0.000	0.001*	–0.003*	–0.052
	(0.001)	(0.000)	(0.001)	(0.001)	(0.028)
Share of households in EA:					
Using nearest school	0.381*	–0.037	–0.011	0.067	–0.002
	(0.033)	(0.023)	(0.041)	(0.049)	(0.033)
With literate head	0.110*	0.001	0.072	–0.076	–0.017
	(0.047)	(0.029)	(0.056)	(0.065)	(0.015)
Per adult equivalent consumption expenditure quintiles (excluded category is quintile 1):[2]					
Quintile 2	0.055*	–0.008	–0.030	0.052	–0.039*
	(0.016)	(0.013)	(0.023)	(0.027)	(0.014)
Quintile 3	0.065*	–0.027*	–0.026	0.071*	–0.042*
	(0.018)	(0.012)	(0.022)	(0.028)	(0.013)
Quintile 4	0.077*	–0.029*	–0.034	0.084*	–0.004
	(0.020)	(0.012)	(0.024)	(0.028)	(0.020)
Quintile 5	0.090*	–0.001	0.000	0.006	–0.007
	(0.024)	(0.017)	(0.030)	(0.036)	(0.004)

Table A.4 (continued)

	Estimated Changes in Probability Associated with Change in Indicated Regressor, from Probit estimates, Various Samples and Dependent Variables[1]				
	Entire Sample	Those who registered for school last year			
	Ever attended school?	Dropped out?	Repeated?	Advanced?	Unclear?
Size of household in adult equivalents	0.006 (0.004)	–0.007* (0.003)	–0.014* (0.006)	0.021* (0.007)	–0.005 (0.027)
Characteristics of household head:					
Male[2]	–0.006 (0.030)	–0.011 (0.026)	0.049 (0.036)	–0.044 (0.044)	–0.006* (0.002)
Age in years	0.005 (0.003)	–0.004 (0.002)	0.004 (0.004)	0.003 (0.004)	0.000* (0.000)
Age squared	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	–0.025 (0.013)
Literate[2]	0.072* (0.016)	–0.028* (0.011)	–0.031 (0.020)	0.058* (0.024)	0.018 (0.022)
Number of spouses	0.012 (0.025)	0.012 (0.021)	–0.019 (0.031)	–0.003 (0.037)	0.031 (0.061)
Share of household members who are:					
Under 7 years old	–0.050 (0.056)	0.022 (0.051)	0.144 (0.081)	–0.172 (0.097)	–0.001 (0.063)
Male and over 15 years	–0.103 (0.062)	–0.003 (0.053)	0.085 (0.084)	–0.071 (0.104)	0.012 (0.066)
Female and over 15 yrs.	–0.027 (0.066)	–0.030 (0.056)	0.083 (0.089)	–0.093 (0.107)	–0.025* (0.012)
Child characteristics (excluded categories are female and age 7):[2]					
Male	0.119* (0.012)	–0.017 (0.010)	–0.011 (0.017)	0.037 (0.019)	0.005 (0.035)
Age 8	0.142* (0.023)	0.079 (0.069)	–0.012 (0.043)	0.011 (0.054)	–0.013 (0.029)
Age 9	0.244* (0.023)	0.048 (0.055)	–0.054 (0.033)	0.076 (0.042)	–0.007 (0.029)
Age 10	0.322* (0.024)	0.074 (0.061)	–0.087* (0.030)	0.112* (0.040)	0.004 (0.031)
Age 11	0.327* (0.026)	0.077 (0.064)	–0.123* (0.025)	0.146* (0.036)	–0.022 (0.026)
Age 12	0.359* (0.023)	0.063 (0.055)	–0.133* (0.028)	0.175* (0.036)	0.006 (0.030)
Age 13	0.361* (0.025)	0.101 (0.067)	–0.125* (0.025)	0.146* (0.036)	–0.012 (0.028)
Age 14	0.380* (0.024)	0.080 (0.060)	–0.159* (0.022)	0.197* (0.033)	

[1,2] See Table A.1, notes 1 and 2.

* Asterisks identify coefficient estimates that are significantly different from zero at the two-tailed five-percent level.

Table A.5. Estimates of Probability Derivatives Employing the WMS/HICES 2000
Samples: Urban Children 7 to 14 Years Old
Dichotomous Dependent Variables Related to Progress Through School

	Estimated Changes in Probability Associated with Change in Indicated Regressor, from Probit estimates, Various Sub-samples and Dependent Variables[1]				
	Entire Sample	Those who registered for school last year			
	Ever attended school?	Dropped out?	Repeated?	Advanced?	Unclear?
Number of Observations	8,076	6,263	6,263	6,263	6,263
Region of residence (excluded category is Amhara):[2]					
Tigray	0.005	–0.007*	0.094*	–0.097*	–0.004
	(0.019)	(0.003)	(0.036)	(0.038)	(0.009)
Affar	–0.034	0.008	0.048	–0.084*	0.033
	(0.040)	(0.012)	(0.032)	(0.032)	(0.021)
Oromiya	–0.010	0.005	0.009	–0.017	0.007
	(0.017)	(0.005)	(0.013)	(0.016)	(0.008)
Somali	–0.171*	0.014	0.057*	–0.073*	0.014
	(0.041)	(0.012)	(0.027)	(0.031)	(0.016)
Benshangul-Gumuz	–0.002	0.009	0.122*	–0.186*	0.066*
	(0.027)	(0.009)	(0.039)	(0.041)	(0.028)
SNNPR	–0.067*	0.009	0.052*	–0.093*	0.039*
	(0.025)	(0.007)	(0.018)	(0.024)	(0.016)
Gambela	–0.018	0.005	0.019	–0.046	0.032
	(0.033)	(0.008)	(0.027)	(0.035)	(0.018)
Harari	0.057*	–0.002	0.054	–0.068	0.012
	(0.013)	(0.005)	(0.036)	(0.039)	(0.016)
Addis Ababa	0.024	0.000	0.016	–0.036	0.019
	(0.016)	(0.004)	(0.017)	(0.021)	(0.012)
Dire Dawa	–0.103*	0.006	–0.021	0.006	0.015
	(0.043)	(0.008)	(0.017)	(0.026)	(0.018)
Distance in km. to nearest:					
Primary school	0.000	–0.001	–0.001	0.002	–0.001
	(0.003)	(0.001)	(0.002)	(0.003)	(0.002)
Post office	–0.002*	0.000	–0.001*	0.001	0.000
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
All-weather road	–0.003	–0.001	0.002*	–0.002*	–0.004
	(0.003)	(0.002)	(0.001)	(0.001)	(0.003)
Share of households in EA:					
Using nearest school	0.068*	–0.005	–0.016	0.006	0.009
	(0.027)	(0.005)	(0.019)	(0.024)	(0.011)
With literate head	0.156*	0.007	–0.039	0.024	0.015
	(0.027)	(0.005)	(0.022)	(0.026)	(0.011)
Per adult equivalent consumption expenditure quintiles (excluded category is quintile 1):[2]					
Quintile 2	0.017	0.004	–0.021	0.014	0.008
	(0.011)	(0.004)	(0.011)	(0.014)	(0.008)
Quintile 3	0.032*	0.003	–0.016	0.014	0.003
	(0.011)	(0.004)	(0.011)	(0.014)	(0.007)
Quintile 4	0.035*	–0.003	–0.023*	0.026	–0.001
	(0.012)	(0.003)	(0.011)	(0.013)	(0.006)
Quintile 5	0.036*	–0.006*	–0.020	0.030*	–0.009
	(0.012)	(0.003)	(0.012)	(0.013)	(0.006)

Table A.5 (continued)

	Estimated Changes in Probability Associated with Change in Indicated Regressor, from Probit estimates, Various Sub-samples and Dependent Variables[1]				
	Entire Sample	Those who registered for school last year			
	Ever attended school?	Dropped out?	Repeated?	Advanced?	Unclear?
Size of household in adult equivalents	0.005*	–0.001	–0.003	0.005*	–0.001
	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)
Characteristics of household head:					
Male[2]	–0.033*	0.006*	–0.017	0.014	0.004
	(0.013)	(0.003)	(0.014)	(0.017)	(0.006)
Age in years	0.009*	–0.001	0.005*	–0.003	–0.001
	(0.002)	(0.000)	(0.002)	(0.002)	(0.001)
Age squared	0.000*	0.000	0.000*	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Literate[2]	0.083*	–0.001	–0.009	0.001	0.007
	(0.012)	(0.003)	(0.010)	(0.012)	(0.005)
Number of spouses	0.007	–0.005	0.004	0.000	–0.004
	(0.013)	(0.003)	(0.012)	(0.014)	(0.006)
Share of household members who are:					
Under 7 years old	–0.128*	0.040*	0.022	–0.049	0.024
	(0.037)	(0.010)	(0.037)	(0.044)	(0.022)
Male and over 15 years	–0.063	0.008	0.013	–0.012	–0.002
	(0.042)	(0.011)	(0.036)	(0.043)	(0.022)
Female and over 15 yrs.	0.065	0.015	–0.025	0.009	0.015
	(0.038)	(0.010)	(0.032)	(0.039)	(0.021)
Child characteristics (excluded categories are female and age 7):[2]					
Male	0.046*	0.006*	–0.002	–0.008	0.010*
	(0.008)	(0.002)	(0.007)	(0.008)	(0.004)
Age 8	0.064*	–0.004	–0.034*	0.047*	–0.011
	(0.009)	(0.003)	(0.011)	(0.013)	(0.008)
Age 9	0.095*	–0.004	–0.064*	0.081*	–0.010
	(0.007)	(0.003)	(0.008)	(0.011)	(0.007)
Age 10	0.102*	–0.003	–0.067*	0.082*	–0.008
	(0.007)	(0.003)	(0.008)	(0.011)	(0.008)
Age 11	0.104*	0.003	–0.072*	0.085*	–0.004
	(0.007)	(0.006)	(0.007)	(0.010)	(0.009)
Age 12	0.109*	–0.004	–0.081*	0.097*	–0.008
	(0.008)	(0.003)	(0.008)	(0.010)	(0.007)
Age 13	0.088*	0.003	–0.068*	0.077*	0.000
	(0.008)	(0.005)	(0.008)	(0.011)	(0.009)
Age 14	0.087*	0.001	–0.063*	0.081*	–0.011
	(0.009)	(0.005)	(0.009)	(0.011)	(0.007)

[1,2] See Table A.1, notes 1 and 2.

* Asterisks identify coefficient estimates that are significantly different from zero at the two-tailed five-percent level.

Table A.6. Estimates of Probability Derivatives Employing the WMS/HICES 2000
Samples: Children 7 to 14 Years Old
Dichotomous Dependent Variable: Whether Child Ever Attended School

	Estimated Changes in Probability Associated with Change in Indicated Regressor, from Probit estimates, Various Sub-samples[1]			
	Children Aged 7–8	Children Aged 9–10	Children Aged 11–12	Children Aged 13–14
Number of Observations	4,557	4,544	4,212	4,096
Urban residence[2]	0.291*	0.312*	0.300*	0.241*
	(0.032)	(0.028)	(0.026)	(0.028)
Region of residence (excluded category is Amhara):[2]				
Tigray	–0.127*	0.004	0.046	0.058
	(0.036)	(0.041)	(0.040)	(0.040)
Affar	–0.133*	–0.178*	–0.083	0.014
	(0.045)	(0.052)	(0.051)	(0.057)
Oromiya	–0.073*	–0.055	–0.014	–0.003
	(0.031)	(0.031)	(0.027)	(0.029)
Somali	–0.277*	–0.298*	–0.272*	–0.234*
	(0.032)	(0.052)	(0.050)	(0.055)
Benshangul-Gumuz	–0.011	0.084	0.145*	0.094*
	(0.055)	(0.047)	(0.034)	(0.041)
SNNPR	–0.124*	–0.111*	0.036	0.030
	(0.032)	(0.036)	(0.030)	(0.029)
Gambela	0.057	0.168*	0.190*	0.201*
	(0.068)	(0.060)	(0.036)	(0.035)
Harari	0.087	0.131*	0.102*	0.118*
	(0.048)	(0.044)	(0.036)	(0.034)
Addis Ababa	0.123*	0.088*	0.107*	0.048
	(0.044)	(0.036)	(0.032)	(0.034)
Dire Dawa	–0.154*	–0.190*	–0.058	–0.054
	(0.043)	(0.052)	(0.049)	(0.048)
Distance in km. to nearest:				
Primary school	–0.035*	–0.031*	–0.010*	–0.018*
	(0.006)	(0.005)	(0.003)	(0.005)
Post office	–0.001	–0.002*	–0.001	–0.001
	(0.001)	(0.001)	(0.001)	(0.001)
All-weather road	0.000	0.000	–0.001	–0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Share of households in EA:				
Using nearest school	0.231*	0.314*	0.426*	0.286*
	(0.045)	(0.043)	(0.039)	(0.040)
With literate head	0.278*	0.159*	0.179*	0.254*
	(0.055)	(0.057)	(0.052)	(0.050
Per adult equivalent consumption expenditure quintiles (excluded category is quintile 1):[2]				
Quintile 2	0.036	0.060*	0.045*	0.039
	(0.026)	(0.024)	(0.022)	(0.022)
Quintile 3	0.057*	0.062*	0.083*	0.024
	(0.027)	(0.026)	(0.022)	(0.023)
Quintile 4	0.116*	0.074*	0.053*	0.028
	(0.030)	(0.026)	(0.025)	(0.026)
Quintile 5	0.182*	0.069*	0.048	0.017
	(0.032)	(0.028)	(0.026)	(0.026)

Table A.6 (continued)

	Estimated Changes in Probability Associated with Change in Indicated Regressor, from Probit estimates, Various Sub-samples[1]			
	Children Aged 7–8	Children Aged 9–10	Children Aged 11–12	Children Aged 13–14
Size of household in adult equivalents	0.011*	0.005	0.013*	0.004
	(0.005)	(0.005)	(0.005)	(0.005)
Characteristics of household head:				
Male[2]	–0.053	0.010	–0.026	–0.020
	(0.041)	(0.038)	(0.032)	(0.031)
Age in years	0.005	0.009*	0.012*	0.013*
	(0.005)	(0.004)	(0.004)	(0.004)
Age squared	0.000	0.000*	0.000*	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Literate[2]	0.139*	0.135*	0.092*	0.050*
	(0.022)	(0.021)	(0.020)	(0.021)
Number of spouses	–0.019	–0.001	–0.012	0.065*
	(0.035)	(0.032)	(0.028)	(0.028)
Share of household members who are:				
Under 7 years old	–0.043	–0.039	–0.121	–0.240*
	(0.082)	(0.082)	(0.083)	(0.080)
Male and over 15 years	0.048	–0.043	–0.140	–0.255*
	(0.101)	(0.093)	(0.083)	(0.078)
Female and over 15 yrs.	0.214*	0.094	0.058	–0.003
	(0.098)	(0.091)	(0.091)	(0.079)
Child characteristics (excluded categories are female and age 7):[2]				
Male	0.067*	0.095*	0.115*	0.149*
	(0.016)	(0.018)	(0.017)	(0.016)
Age 8	0.152*			
	(0.017)			
Age 9				
		–0.056*		
Age 10		(0.018)		
Age 11			–0.019	
			(0.016)	
Age 12				
Age 13				–0.006
				(0.016)
Age 14				

1,2 See Table A.1, notes 1 and 2.

* Asterisks identify coefficient estimates that are significantly different from zero at the two-tailed five-percent level.

Table A.7. Estimates of Probability Derivatives Employing the WMS/HICES 2000
Samples: Rural Children 7 to 14 Years Old
Dichotomous Dependent Variable: Whether Child is Currently Registered for School

	Estimated Changes in Probability Associated with Change in Indicated Regressor, from Probit estimates, Various Sub-samples[1]					
	2 km. or less from nearest primary school	More than 2 km. from nearest primary school	Per adult equivalent expenditure quntiles 1 and 2	Per adult equivalent expenditure quintiles 3, 4, and 5	Boys	Girls
Number of Observations	5,900	7,103	7,122	5,881	4,813	4,520
Region of residence (excluded category is Amhara):[2]						
Tigray	0.077	0.035	0.021	0.108*	–0.034	0.016
	(0.045)	(0.035)	(0.033)	(0.053)	(0.043)	(0.033)
Affar	–0.113*	–0.142*	–0.090*	–0.216*	–0.077	–0.144*
	(0.049)	(0.041)	(0.043)	(0.028)	(0.071)	(0.034)
Oromiya	–0.062*	–0.008	–0.046*	–0.015	0.043	–0.148*
	(0.029)	(0.021)	(0.022)	(0.024)	(0.032)	(0.020)
Somali	–0.156*	–0.188*	–0.203*	–0.123*	–0.148*	–0.195*
	(0.044)	(0.016)	(0.032)	(0.047)	(0.036)	(0.023)
Benshangul-Gumuz	0.177*	0.159*	0.122*	0.285*	0.300*	–0.030
	(0.048)	(0.036)	(0.037)	(0.053)	(0.046)	(0.039)
SNNPR	–0.020	0.057*	–0.004	0.050	0.072*	–0.118*
	(0.030)	(0.023)	(0.022)	(0.030)	(0.033)	(0.021)
Gambela	0.274*	0.427*	0.326*	0.346*	0.423*	0.199*
	(0.055)	(0.053)	(0.049)	(0.068)	(0.045)	(0.070)
Harari	0.053	0.058	–0.003	0.097*	0.241*	–0.069*
	(0.040)	(0.042)	(0.044)	(0.037)	(0.053)	(0.032)
Addis Ababa	0.145	0.056	0.103*	0.075	0.082	0.093*
	(0.085)	(0.037)	(0.048)	(0.051)	(0.050)	(0.048)
Dire Dawa	–0.060	–0.081*	–0.109*	–0.004	–0.033	–0.158*
	(0.053)	(0.026)	(0.027)	(0.043)	(0.044)	(0.027)
Distance in km. to nearest:						
Primary school	–0.052*	–0.014*	–0.016*	–0.023*	–0.021*	–0.024*
	(0.011)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
Post office	0.000	0.000	0.000	0.000	–0.000	–0.001
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
All-weather road	–0.001	–0.002*	–0.002*	–0.001	–0.001	–0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of households in EA:						
Using nearest school	0.372*	0.303*	0.347*	0.376*	0.393*	0.341*
	(0.043)	(0.034)	(0.035)	(0.041)	(0.041)	(0.038)
With literate head	0.141*	0.055	0.146*	0.045	0.104	0.091
	(0.061)	(0.045)	(0.045)	(0.056)	(0.060)	(0.053)
Per adult equivalent consumption expenditure quintiles (excluded category is quintile 1):[2]						
Quintile 2	0.077*	0.034*	0.051*		0.073*	0.034
	(0.021)	(0.017)	(0.014)		(0.021)	(0.019)
Quintile 3	0.069*	0.044*			0.093*	0.041
	(0.025)	(0.019)			(0.024)	(0.024)
Quintile 4	0.073*	0.043*		–0.009	0.110*	0.060*
	(0.026)	(0.020)		(0.019)	(0.027)	(0.024)

Table A.7 (continued)

	Estimated Changes in Probability Associated with Change in Indicated Regressor, from Probit estimates, Various Sub-samples[1]					
	2 km. or less from nearest primary school	More than 2 km. from nearest primary school	Per adult equivalent expenditure quntiles 1 and 2	Per adult equivalent expenditure quintiles 3, 4, and 5	Boys	Girls
Quintile 5	0.101*	0.032		–0.005	0.084*	0.082*
	(0.030)	(0.024)		(0.018)	(0.032)	(0.029)
Size of household in adult equivalents	0.018*	0.012*	0.013*	0.020*	0.009	0.004
	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Characteristics of household head:						
Male[2]	–0.082*	–0.021	–0.062*	–0.025	–0.029	0.017
	(0.037)	(0.031)	(0.030)	(0.042)	(0.041)	(0.033)
Age in years	0.013*	0.005	0.006	0.012*	0.005	0.007*
	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Age 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)
Literate[2]	0.089*	0.064*	0.051*	0.099*	0.067*	0.081*
	(0.021)	(0.017)	(0.019)	(0.019)	(0.022)	(0.021)
Number of spouses	0.028	–0.002	0.040	–0.035	0.025	–0.010
	(0.032)	(0.026)	(0.025)	(0.035)	(0.033)	(0.029)
Share of household members who are:						
Under 7 years old	–0.066	–0.037	–0.053	–0.039	–0.048	–0.030
	(0.075)	(0.055)	(0.062)	(0.066)	(0.074)	(0.070)
Male and over 15 years	–0.014	0.058	0.144*	–0.059	–0.104	–0.026
	(0.070)	(0.053)	(0.057)	(0.063)	(0.086)	(0.072)
Female and over 15 yrs.	0.013	0.006	0.047	–0.015	–0.134	0.096
	(0.078)	(0.056)	(0.062)	(0.073)	(0.091)	(0.079)
Child characteristics (excluded categories are female and age 7):[2]						
Male	0.161*	0.119*	0.140*	0.147*		
	(0.016)	(0.011)	(0.012)	(0.015)		
Age 8	0.016	–0.039*	–0.069*	0.045	0.138*	0.115*
	(0.026)	(0.018)	(0.021)	(0.025)	(0.032)	(0.029)
Age 9	0.144*	0.007	0.032	0.116*	0.274*	0.172*
	(0.027)	(0.020)	(0.023)	(0.026)	(0.031)	(0.032)
Age 10	0.179*	0.091*	0.114*	0.162*	0.309*	0.282*
	(0.025)	(0.024)	(0.025)	(0.027)	(0.032)	(0.034)
Age 11	0.146*	0.115*	0.104*	0.173*	0.357*	0.234*
	(0.028)	(0.024)	(0.026)	(0.029)	(0.032)	(0.036)
Age 12	0.179*	0.142*	0.142*	0.194*	0.398*	0.248*
	(0.025)	(0.021)	(0.023)	(0.025)	(0.029)	(0.033)
Age 13	0.213*	0.084*	0.137*	0.166*	0.400*	0.228*
	(0.028)	(0.023)	(0.024)	(0.031)	(0.031)	(0.037)
Age 14	0.206*	0.132*	0.159*	0.192*	0.399*	0.280*
	(0.027)	(0.025)	(0.025)	(0.027)	(0.030)	(0.035)

[1,2] See Table A.1, notes 1 and 2.

* Asterisks identify coefficient estimates that are significantly different from zero at the two-tailed five-percent level.

Table A.8. Answers to Question: "Why doesn't anyone in your household use the nearest primary school?" from the WMS/HICES 2000
Sample: Children ages 7 to 14 in households responding that they do not use the nearest primary school

	Percentage distribution of responses	
Response	Rural Areas	Urban Areas
"Too expensive"	0.5	11.2
"Too far"	7.6	0.7
"Poor quality service"	0.6	11.4
"Incomplete service"	0.1	3.8
"Have no experience"	13.6	6.9
"Have no need of it"	26.5	36.7
"Other/specify"	51.2	29.3

Analysis of the Merged LFS/EMIS 1999 Data

This appendix presents some basic results regarding the determinants of whether children have ever attended school, and other schooling and work outcomes, for the merged LFS 1999–EMIS (school census) 1999 data. While the questionnaire for the LFS 1999 is not as rich as those for the other datasets employed in the larger project, it employs the largest sample, and contains sufficient information that it may be matched with school census data at the woreda level.

Matching the LFS to School Census Data. The geographic codes for the LFS and the school census were not identical. Thus the matching was done on the basis of woreda names. While the spellings were dramatically different across the two datasets, it was possible (working "by hand") to obtain matches in the school census data for 425 of the 461 woredas represented in the LFS data. Because it was necessary to express some of the school census information on a per capita (within woreda) basis, I also merged in population totals by woreda from the Statistical Abstract 2001. The spelling and ordering of the woreda names in this document were almost identical to those in the LFS 1999. The document also contained information on population density, which serves as a useful control for heterogeneity in economic development and urbanization across woredas. Unfortunately, the density information was missing for another 15 woredas, all in Affar and Somali, causing those two regions (which are underrepresented in the first place) to fall out of the analysis.

Samples. In most of what follows I restrict attention to children ages 7 to 14 for whom all the relevant variables are non-missing. While the main dependent variable (whether the child has ever attended school) is available for 84,345 children, the sample size falls to 84,299 when missing values for other LFS variables are dropped. More seriously, the sample size falls to 78,941 when observations for which school census and woreda population information could not be merged are dropped. Finally, the sample size falls to 74,093 when observations for which the woreda population density variables are missing are dropped.

Variable Definitions and Descriptive Statistics. The first dependent variable employed below is an indicator of whether or not the child has ever attended school. Three additional dependent variables have to do with current time use:

whether the child is currently attending school, whether the child is engaged in "generating income" for the household (whether through wage employment or work in a household agricultural or non-agricultural enterprise) and whether the child participates in unpaid housework. The latter two variables are derived from two questions. 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. A child is coded as involved in generating income if the response to either question was coded as 1, 2, 3, 4, 5 or 6. The child is coded as engaged in unpaid housework if the response was coded as 7. As is apparent in the table, many children work, either in income generation (primarily on family farms) or unpaid housework. Income generation is more prevalent in rural than urban areas, while the unpaid housework is more prevalent in urban areas.

The first two columns of Table B.1 present descriptive statistics, within rural and urban sub-samples, for the main potential determinants employed in the analysis.

Region Indicators. "Region" indicators are included to capture systematic differences across regions in school supply and demand factors that are not captured explicitly by other variables included in the regressions. While the remaining controls allow for relatively rich treatment of school supply differences that might differ across regions, they provide less complete control for differences in economic circumstances and in cultural factors. Thus it is these latter effects that are most likely to be picked up by the region indicators in this application.

Woreda School Availability, Woreda Density, Woreda Teachers, and Woreda School "Quality." The basic indicators of school availability at the woreda level are the numbers of schools in three categories per 1000 population within the woreda. The three categories are schools that have classes in the grades 1–4 range, in the grades 5–8 range and in the grades 9–12 range. These are not mutually exclusive categories. They are expressed on a per 1000 population basis, for the obvious reason that 10 schools implies much better availability in a small woreda than in a large woreda. This is, of course, an incomplete description of school availability. Consider two woredas of the same total population and containing the same number of schools, and let the schools be evenly spread out within the geographic areas of the woredas. If the woredas differ in population density (in, say, population per square km.) the typical child will live closer to a school in the more densely populated woreda. Thus it is useful to control not only for schools per 1000 population, but also for population density. The distribution of woredas by density is skewed to the right, and the relationship between schooling rates and density appears to be quite nonlinear (results not shown). Thus both density and its square are included, and it may also be useful to assess robustness of results to the elimination of unusual values of the density variable.

The supply of schooling depends not only on how many schools there are but also how many classes and students may be accommodated per school. I attempt to capture this with the three variables, capturing the numbers of teachers in each of the three grade ranges divided by the numbers of schools with classes in each of these grade ranges. (In woredas in which there are no schools in a particular grade range, this teacher quantity variable is also set to zero.)

The remaining school variables represent an attempt to capture variation in typical school quality across woredas. They include the percentages of schools in the grade 1–8 range that have latrines and libraries, the percentage of grade 1–8 teachers who are female, the percentage of grade 1–8 teachers who have “less than TTI qualification” (meaning they have only 6–12 years of schooling and no teacher training certificate), and the percentage of grade 1–8 teachers who were reported to have vocational training, specialized training in physical education, crafts, agriculture, music or art, or BA or MA degrees.

EA Average Variables. Some specifications below include the EA average variables. These variables are calculated from LFS sample data. They make use of the data on the heads of all households in the same enumeration area except the household of the child to which the observation pertains. (The child’s own household is excluded from the calculation of this average, to avoid introducing spurious correlation between this right hand side variable and the dependent variable.) The shares of household heads who are literate, have formal jobs (defined below) and have no jobs are included to capture variation across communities in economic conditions and in exposure to schooling.

Household Head Variables. The next set of variables describes the household head. In the LFS we cannot identify the child’s own parents. The best we can do is to identify the head of the household to which the child belongs. The household head variables serve at least two functions. First, they come the closest we can get to capturing variation in the households’ level of economic resources. Unfortunately, the LFS data contain no direct measures of income, consumption expenditure or assets. Thus the education level, age and employment status of the household head are the best indication we have of how well off the household is. By “formal job” is meant a permanent or contract job in wage employment, whether in the public or private sector, or an NGO. Data from many other countries tend to show that such jobs pay better than other jobs. Second, the household head variables capture variation across households in preferences, related to the education level of the head (here, an indicator of whether the head is “literate”, which is taken by the survey designers to mean that s/he has had some schooling), the head’s religion, and whether the head has ever migrated.

Household Structure. Household structure (e.g. the fraction of household members who are young children, and who are men and women over 14 years of age) may influence the opportunity cost of a school-age child’s time. More young children means more child care needs within the household. More adult females may mean greater potential to replace school-aged children by adults as child-care givers, while more adult males may increase the potential productivity of child labor in household enterprises. As an additional measure of household resources, I include the share of household members with “formal” jobs.

Having controlled for whether the head has a formal job and the total share of members with formal jobs, household size may enter for a couple of reasons, rendering its coefficient difficult to interpret. If the other variables proxy well for income per capita, then it might take a positive coefficient, picking up economies of scale. If however, the other variables proxy better for total income than for income per capita, it might take a negative sign, as it indicates the need to stretch the same income over more needs.

Child Variables. Among the “child” variables are indicators of the child’s sex and age, as well as indicators of whether the child’s mother and father are not reported as alive. (That is, this variable equals one whether the response to the question of whether the parent is alive is “no”, "don't know" or "missing.") The indicator that both parents are "not alive" is included in case the effects of losing mother and father are not additive.

Difficulties in Interpreting the School Supply Effects. The main attraction of working with this dataset is the ability to merge household-level data on schooling choices with administrative data on the availability and quality of schools by region. Unfortunately, it is important to point out several difficulties with estimating and interpreting the school supply variable effects in the regressions below. First, the school supply variables themselves might contain inaccuracies. See Schaffner and Banki (2002) for a detailed description of the administrative data employed.

Second, the school supply data are aggregated to a geographic level that is much larger than would be ideal for this study. Woredas are geographic regions that are contained within administrative regions (killils), but that span both rural and urban sub-areas. They are larger than enumeration areas, or the "communities" we might like to observe. Given this level of aggregation, the introduction of school supply controls may help to explain differences in schooling rates across larger administrative regions (killils), but is unlikely to help explain variation between rural and urban areas. Even if school supply is dramatically different between rural and urban areas, and even if this difference in supply plays a large role in explaining overall rural-urban differences, we are unlikely to pick up such effects in these data. [Notice that, in the descriptive statistics presented separately for rural and urban sub-samples of the LFS in the first two columns of Table B.1, the woreda-level variables derived from school and population census data do not differ greatly across the sub-samples. This apparent similarity in supply is an artifact of the level of aggregation.] Note furthermore that this high level of aggregation implies measurement error of a sort that might be expected to bias coefficients on school supply variables toward zero. School supply variables aggregated to the woreda level represent the average of comparable measures at the community level. Woreda-level supply variables can be thought of as measuring true community-level supply variables with errors that average to something near zero within woredas. The attenuation bias associated with classical measurement error would suggest that we will underestimate the effect of, say, schools per 1000 population on registration rates.

Third, the final data set contains only 424 woredas, and the school supply variables at the woreda level tend to be correlated. Thus multicollinearity is likely to render it difficult to get precise estimates of individual school supply characteristic effects.

Finally, potential problems of endogeneity or omitted relevant variables render the coefficients on the school supply variables difficult to interpret, even if precisely estimated. Intuitively, we suspect that choices regarding school building and equipping have been driven in part by community characteristics that we do not observe, and these unobserved characteristics probably help determine enrollment rates. Under such circumstances, the coefficients on school supply variables pick up not only the intrinsic effect of expanding school supply, but also the effect of these unobserved community characteristics. If schools have tended to be built in communities that enjoy higher levels of development and of interest in the schooling of children, then our coefficients will tend to over-estimate the effect of expanding school supply while holding level of development constant. On the other hand, if schools have tended to be built in more disadvantaged areas, where the disadvantage is a type that tends to inhibit school enrollment (even after the schools are built), then the coefficient on school supply will tend to understate the effect of expanding school supply while holding these advantages and disadvantages constant.

Basic Multivariate Results Regarding the Determinants of "Ever Attendance". Tables B.1 through B.4 report estimates of coefficients in a variety of specifications relating the probability that a child has ever attended school to some or all of the potential determinants. Estimates in columns 3 through 7 of Table B.1 employ pooled (urban plus rural) samples, while the specifications in Tables B.2, B.3 and B.4 pertain to rural and urban sub-samples.

The main observations from Table B.1 are the following:

- Rural-urban differences remain large, even in the presence of all the controls possible in this dataset. This is not very surprising, as the dataset contains very little indication of household income or asset levels, and as the school supply variables are observed only at a level of aggregation that crosses rural and urban boundaries.

- Many of the regional differences seen in the third column of Table B.1 are, however, reduced or even reversed after the inclusion of the school supply variables. This suggests that school supply differences plays an important role in explaining school registration differences across regions. One suspects that they would play a role in explaining rural–urban differences, too, if the supply measures were aggregated over smaller geographic units, allowing differentiation between rural and urban areas.

- Instability of coefficients and large standard errors suggest that the school supply variables are quite collinear. Coefficients on the variables describing numbers of schools per 1000 population, however, give the impression that registration rates are higher in woredas with more schools per person. A coefficient of .29 on the number of schools with grades 1–4 per 1000 population, for example, indicates that an increase of .2 in schools per 1000 population would increase ever attendance rates by about .2*.29=5.8 percent. (Primary school aged children constitute about 1/5 of the population. Thus increasing schools by .2 per 1000 is the same as increasing schools by .2 per 200, or 1 per 1000 children.)

- There even seems to be something about woredas in which there are more secondary schools per 1000 population that tends to produce higher registration rates in primary schooling. It is difficult to pin the estimated effect on secondary school supply per se, however.

- There is little evidence of large effects of increases in teachers per school on registration rates, holding the number of schools and population density constant.

- The effects of school quality measures are imprecisely estimated, but jointly significant at least at the 10 percent significance level. While it is difficult to distinguish the particular features of school supply that are of greatest importance for attendance rates, it seems clear that attendance rates are higher in woredas that have invested more in creating higher quality school systems.

- Holding all else equal, children who live in communities in which more of the household heads are literate are much more likely to have attended school. This bolsters the view that community exposure to school and attitudes about schooling might play an important role in determining schooling rates.

- School attendance rates are also higher in communities where more of the household heads are unemployed, suggesting that opportunity costs of children's time might influence schooling choices, and that such opportunity costs are lower where labor markets are more slack.

- Even controlling for literacy rates among other household heads in the community, the literacy of the child's own household head has a large positive effect on the probability of school attendance, though the estimated effect falls as better and better controls are introduced for community level attitudes and exposure to schooling.

- In general children of household heads that have ever migrated are more likely to have attended school, but within communities (enumeration area fixed effects estimates) these differences disappear. This indicates that the overall difference between children of migrants and non-migrants is driven in great part by differences in their geographical location.

- Children of heads with formal jobs, who probably enjoy higher standards of living, are also more likely to have attended school. Again, a big part of the overall difference seems to be explained by differences in the geographic location of children with heads who have different kinds of jobs. Even within communities, however, attendance rates are perhaps 5 percentage points higher for children of heads with formal jobs.

- After controlling for community unemployment percentages, children of heads who have no job have attendance rates little different from children whose heads have non-formal jobs.

- Larger numbers of younger siblings or of adult males both seem to reduce attendance rates.

- Children from households in which a larger fraction have formal jobs are more likely to attend school, but much of this effect disappears in the enumeration fixed effects estimates. This suggests that it is picking up differences across geographic areas in the nature of labor markets, rather than household resource effects.

- Male children are significantly more likely to attend school than female children.

- Having lost a parent puts a child at a disadvantage regarding school attendance (of about 5 percentage points per parent), and the effects of losing parents appears to be roughly additive.

The main observations from Table B.2 are:

- The apparent effects of increasing numbers of school per 1000 population are larger in rural than urban areas. Given the way in which the school supply variables are defined (aggregated to the woreda level), this is not just capturing a nonlinearity in the school supply effect (larger absolute effects where school supply rates are lower). In fact, as seen in Table B.1, average school supply measures are higher in the rural sample than in the urban sample. What the difference may be capturing is that woreda-level variation in school supply variables makes more of a difference to the rural parts of woredas than to the urban parts of the woredas. Given the high share of the population living in rural areas, woreda averages are also probably subject to less severe measurement errors in the rural sub-sample than in the urban sub-sample.

- Woreda-level population density variables matter more for attendance rates in the rural sample than in the urban sample. Among rural areas, though that are parts of more urban, and thus more densely populated, woredas exhibit higher attendance rates. Perhaps being part of a more urban woreda gives the woreda more exposure to labor markets in which schooling is beneficial.

- School infrastructure variables (shares of schools having latrines and libraries) seem to have stronger associations with attendance rates in urban than rural areas.

- The effects of community characteristics (as proxied by enumeration area average variables) have stronger effects in rural than urban areas.

- The presence of younger siblings has a stronger negative effect on school attendance in urban areas than rural areas, while the presence of more adult males has a stronger negative effect on school attendance in rural areas than urban areas. The forces shaping the opportunity cost of children's time appears to be rather different between rural and urban areas, as is no surprise. Having more adult males on a farm may create additional profitable opportunities for children to work and learn by doing in agriculture. Younger siblings may be a stronger drag on school attendance in urban areas, if mothers are more likely to work in jobs to which they cannot bring their young children.

- The share of the household with formal jobs (which are more unusual in rural areas) has a larger effect on enrollment rates in the rural sample than in the urban sample. In part this probably reflects the greater effect of geographic differences in labor markets within rural areas (because the effect diminishes in the enumeration area fixed effects specification).

- Gender differences in schooling rates are higher in rural than urban areas.

- The school enrollment disadvantage of losing parents appears similar in rural and urban areas.

The main observations from Table B.3 are:

- Within the rural sample, regional differences in registration rates appear quite different for boys and girls, indicating differences in the nature and importance of gender concerns across regions.

- Numbers of schools per 1000 population and population density have larger effects on attendance by rural boys than by rural girls. This contributes to a guess that differences in attendance rates between rural boys and girls will tend to widen as school supply improves and as economic development proceeds, at least at first.

- Within rural areas, most household head variables have very similar effects on attendance rates of boys and girls. The exception is the indicator of whether the head has a formal job, which plays a larger role in determining boys' attendance rates than girls. In urban areas, household head variables seem to matter more for girls' rates than for boys'. The meaning of this difference is unclear, as the coefficient on the share of household members who have formal jobs is much bigger for girls than for boys.

- In both rural and urban areas, larger numbers of younger siblings has a bigger downward pull on attendance rates for girls than boys, consistent with the view that girls are considered more responsible for child care than are boys.

- The presence of more adult males pulls down school attendance rates for both girls and boys in rural areas.

The main observation from Table B.4 is:

- In the household fixed effects specification, the differences in school attendance rates between children who have and have not lost parents are even greater than in the other specifications. This indicates that the lower attendance rates are not just the result of the presence of orphaned relatives in a household driving down the level of resources per person. Even within households (and thus holding household resource levels constant), the children who have lost parents are less likely to attend school.

Current Time Use

Rather than examining the determinants of whether children have ever attended school, we may examine the determinants of how they are currently using their time, using three dichotomous dependent variables indicating whether or not the child is attending school, whether or not the child is engaged in income generation and whether or not the child is engaged in unpaid housework. Questionnaire design (described above) implies that a child may report only one of the two work activities: income generation and unpaid housework. The child may, however, report both attending school and being engaged in one of the two work activities.

Table B.5 presents some simple cross-tabulations on participation rates in these activities, separately for rural and urban boys and girls. Differences in rates of current school attendance mirror differences in the rates of "ever attendance" studied above, because the great majority of children who have ever attended are currently in school. Rates of participation in income generation vary greatly across groups, being highest for rural boys and lowest for urban girls, as one might guess. Rates of participation in unpaid housework are higher for girls than for boys, and higher in urban than rural areas.

Among rural boys, while rates of participation in jobs are higher for boys who are not in school than for those who are in school, the difference is small, and even 51 percent of rural boys who are in school are also working in a job. Thus much rural child labor, at least of the sort that boys may engage in, appears to be compatible with school attendance. The differences in job participation rates across children who are and are not in school are greater for rural girls and for all urban children.

For all four groups examined in the table, rates of reported participation in unpaid housework are higher among children who are attending school than among those who are not attending school. Thus the household work being captured by this variable contains at least some important components that are compatible with school attendance. It is possible that this greater involvement in household work helps to compensate for reductions in child earnings from jobs or for the cost of schooling fees, by freeing up the time of adult members of the household to be outside the home earning income. At least in part, however, it probably reflects the fact that even children who are involved in income generation are also involved in unpaid housework. Because the questionnaire elicits participation in income generation and housework in a mutually exclusive fashion, involvement in house work only "shows up" for children who are not involved in income generation.

Tables B.6 and B.7 present the results of probit estimations in which the three time use dependent variables are related to the determinants of school attendance employed above, first for rural and urban boys, and then for rural and urban girls. We are interested in observing whether the same economic factors that increase schooling also decrease participation in child labor, suggesting that child labor and schooling are competing uses for child time. If, instead, economic factors increase school attendance without diminishing child labor, this would suggest that labor and schooling are potentially complementary, and that the opportunity cost of children's time (associated with involvement in jobs) is not a large force inhibiting school attendance.

The main observations from Tables B.6 and B.7 are the following:

- Many of the estimates are statistically insignificant.

- Some school supply factors that have strong estimated effects on school attendance rates do not seem to have important effects on work and household rates.

- Enumeration area rates of literacy among household heads have large positive effects on school attendance, but smaller negative effects on job participation rates and smaller positive effects on household participation rates in rural areas.

- Where a higher share of household heads are engaged in formal jobs, children have higher school attendance rates and lower job participation rates, in both rural and urban areas, though the size of the effects is larger in rural areas.

- Where rates of unemployment among household heads are higher, school attendance rates are higher, while child job participation rates are much lower and child housework participation rates are higher.

- Children with household heads who are literate are much more likely to attend school and somewhat more likely to do unpaid housework, but are little different from others in their propensity to work at jobs.

- Children whose household heads have formal jobs are substantially more likely to attend school, and, for rural girls, more likely to do housework and less likely to work at jobs.

- As the share of young dependents within the household rises, school attendance rates fall (more so in urban than rural areas), job participation rates rise (more so in rural than urban areas), and participation in unpaid household work rises (more so in urban than rural areas).

- As the share of adult males within the household rises, school attendance rates fall and participation in jobs rises among rural children.

- As the share of adult females within the household rises, children's participation in unpaid housework falls, especially for girls. A corresponding significant increase in school participation is picked up only for urban boys.

- School attendance rates are significantly lower for children who have lost parents. Their participation in work activities is significantly higher only among rural boys.

Subsidiary Information

The LFS offers a little information on the motivations for migration. In particular, it asks of people who migrated in the last five years the reason for which they left their previous residence. The possible answers are as given in Table B.8. I tabulate the responses separately for rural and urban areas, and report responses among two groups: children ages 7–14 and the heads of these children's households. Among migrants in urban areas, education is a significant motivation for migration, perhaps more so among children who have migrated without their families than among migrants that move as families.

Table B.1. Descriptive Statistics and Estimates of Probability Derivatives Employing the LFS/EMIS 1999 Samples: Children 7 to 14 Years Old
Dichotomous Dependent Variable: Whether Child has Ever Attended School

			Estimated Changes in Probability Associated with Change in Indicated Regressor, Various Methods[1]				
	Rural Means (Std. Dev.)	Urban Means (Std. Dev.)	Probit	Probit	Probit	OLS	EA Fixed Effects
Number of Observations	49,554	24,539	74,093	74,093	74,093	74,093	74,093
Urban residence[2]			0.475*	0.406*	0.402*	0.292*	
			(0.008)	(0.010)	(0.010)	(0.015)	
Region of residence (excluded category is Amhara):[2]							
Tigray	0.063	0.065	0.043*	0.002	0.026	0.033	
			(0.020)	(0.020)	(0.022)	(0.022)	
Oromiya	0.421	0.338	0.036*	0.015	–0.002	–0.013	
			(0.014)	(0.014)	(0.017)	(0.017)	
Benshangul-Gumuz	0.013	0.007	0.165*	0.061	0.046	0.033	
			(0.022)	(0.035)	(0.035)	(0.036)	
SNNPR	0.241	0.134	0.010	–0.004	–0.029	–0.040*	
			(0.015)	(0.015)	(0.017)	(0.017)	
Gambela	0.002	0.006	0.257*	0.036	0.050	0.031	
			(0.037)	(0.073)	(0.071)	(0.076)	
Harare	0.001	0.010	0.158*	0.095*	0.006	0.019	
			(0.030)	(0.034)	(0.042)	(0.043)	
Addis Ababa	0.001	0.300	0.279*	0.167*	0.141*	0.092	
			(0.019)	(0.037)	(0.045)	(0.048)	
Dire Dawa	0.002	0.003	0.065	0.012	–0.005	0.101	
			(0.038)	(0.060)	(0.058)	(0.053)	
Number of schools in woreda per 1000 population that offer:							
Grades 1–4	0.184(0.074)	0.159(0.078)		0.083	0.294*	0.296*	
				(0.102)	(0.112)	(0.112)	
Grades 5–8	0.110(0.053)	0.108(0.050)		0.387*	0.198	0.072	
				(0.120)	(0.135)	(0.134)	
Grades 9–12	0.005(0.006)	0.011(0.011)		2.183*	0.433	0.638	
				(0.730)	(0.760)	(0.781)	
Woreda density in 1000 population per sq. km.	.202 (.585)	5.203 (12.624)		0.003	–0.001	–0.005	
				(0.004)	(0.004)	(0.004)	
Woreda density squared				–0.002	0.005	0.014*	
				(0.007)	(0.007)	(0.006)	
Number per school in woreda of teachers for:							
Grades 1–4	6.507(1.912)	9.797(4.616)			0.003	0.002	
					(0.003)	(0.003)	
Grades 5–8	5.083(2.516)	9.967(6.263)			0.000	–0.001	
					(0.003)	(0.003)	
Grades 9–12	20.742(20.002)	34.014(26.921)			0.002*	0.002*	
					(0.000)	(0.000)	
Shares of school in woreda with:							
Latrines	0.430(0.156)	0.440(0.134)			0.066	0.060	
					(0.037)	(0.036)	
Libraries	0.182(0.124)	0.209(0.127)			0.101*	0.085	
					(0.048)	(0.048)	
Shares of woreda teachers in grades 1–8 who:							
Are female	0.276(0.088)	0.330(0.096)			–0.092	0.000	
					(0.073)	(0.071)	
Have less than TTI qualification	0.024(0.040)	0.040(0.048)			–0.091	–0.131	
					(0.124)	(0.116)	
Have vocational training, BA or MA	0.002(0.005)	0.017(0.030)			0.537	0.340	
					(0.638)	(0.634)	
Share of households in EA:							
With literate head	0.257(0.131)	0.631(0.174)				0.478*	
						(0.040)	
Whose head has formal sector job	0.010(0.027)	0.216(0.148)				0.031	
						(0.069)	

Table B.1 (continued)

	Rural Means (Std. Dev.)	Urban Means (Std. Dev.)	Estimated Changes in Probability Associated with Change in Indicated Regressor, Various Methods[1] Probit	Probit	Probit	OLS	EA Fixed Effects
Whose head has no job	0.122(0.111)	0.237(0.128)				0.121* (0.048)	
Characteristics of household head:							
Male[2]	0.813	0.679		–0.087* (0.009)	–0.085* (0.009)	–0.076* (0.009)	–0.043* (0.005)
Age in years	46.220(12.431)	44.815(12.571)		0.005* (0.001)	0.005* (0.001)	0.004* (0.001)	0.003* (0.001)
Age squared				–0.003* (0.001)	–0.003* (0.001)	–0.003* (0.001)	–0.003* (0.001)
Literate[2]	0.250	0.629		0.159* (0.007)	0.156* (0.007)	0.116* (0.006)	0.084* (0.004)
Reports traditional religion[2]	0.040	0.004		–0.167* (0.018)	–0.147* (0.018)	–0.117* (0.018)	–0.066* (0.010)
Ever migrated[2]	0.290	0.849		0.058* (0.008)	0.057* (0.008)	0.048* (0.008)	0.007 (0.004)
Has formal job[2]	0.010	0.217		0.134* (0.020)	0.129* (0.020)	0.109* (0.020)	0.053* (0.010)
Has no job[2]	0.092	0.201		0.003 (0.010)	0.003 (0.010)	–0.013 (0.009)	0.003 (0.005)
Household size (number of members)	6.498(2.123)	6.578(2.593)		0.011* (0.002)	0.011* (0.002)	0.010* (0.001)	0.007* (0.001)
Share of household members who are:							
Under 7 years old	0.202(0.152)	0.138(0.142)		–0.129* (0.023)	–0.133* (0.023)	–0.120* (0.023)	–0.053* (0.014)
Male and over 15 years	0.212(0.125)	0.211(0.144)		–0.099* (0.027)	–0.109* (0.027)	–0.121* (0.027)	–0.064* (0.016)
Female and over 15 yrs.	0.224(0.108)	0.283(0.141)		0.066* (0.028)	0.051 (0.028)	0.043 (0.028)	0.010 (0.016)
In formal jobs	0.004(0.028)	0.059(0.110)		0.251* (0.083)	0.245* (0.083)	0.178* (0.083)	0.055 (0.039)
Child characteristics (excluded categories are female and age 7):[2]							
Male	0.514	0.481		0.147* (0.005)	0.148* (0.005)	0.150* (0.005)	0.111* (0.003)
Age 8	0.151	0.123		0.119* (0.008)	0.120* (0.008)	0.120* (0.008)	0.089*s (0.005)
Age 9	0.131	0.116		0.214* (0.007)	0.215* (0.007)	0.217* (0.007)	0.165* (0.006)
Age 10	0.131	0.130		0.278* (0.007)	0.279* (0.007)	0.280* (0.007)	0.228* (0.006)
Age 11	0.085	0.099		0.309* (0.007)	0.309* (0.007)	0.311* (0.007)	0.262* (0.006)
Age 12	0.135	0.140		0.328* (0.007)	0.330* (0.007)	0.331* (0.007)	0.279* (0.005)
Age 13	0.104	0.130		0.335* (0.007)	0.336* (0.007)	0.337* (0.007)	0.289* (0.006)
Age 14	0.102	0.142		0.339* (0.007)	0.339* (0.007)	0.339* (0.007)	0.293* (0.006)
Whether child reports:							
Mother not living[2]	0.056	0.070					
				–0.060*	–0.061*	–0.061*	–0.042*
Father not living[2]	0.120	0.162		(0.012) –0.066*	(0.012) –0.066*	(0.012) –0.064*	(0.007) –0.045*
Both parents not living[2]	0.012	0.026		(0.009) –0.011	(0.009) –0.012	(0.009) –0.018	(0.005) –0.002
				(0.024)	(0.024)	(0.024)	(0.014)

[1] For continuous regressors, the estimates are of the derivative of the probability of the dichotomous dependent variable equalling one with respect to the regressor, evaluated at the means of all right hand side variables. For dichotomous regressors, the estimates are of the change in probability as the variable is changed from zero to one, while holding all other right hand side variables at their means. Robust standard errors are shown in parentheses.

[2] These regressors are dichotomous, taking the value 1 if the indicated condition is true, and zero otherwise. They capture differences in the dependent variable between the indicated category and the excluded category. Where no excluded category is explicitly mentioned in the table, it is the opposite of the category indicated.

* Asterisks identify coefficient estimates that are significantly different from zero at the two-tailed five percent level.

Table B.2. Estimates of Probability Derivatives Employing the LFS/EMIS 1999
Samples: Rural and Urban Children 7 to 14 Years Old
Dichotomous Dependent Variable: Whether Child has Ever Attended School

	Estimated Changes in Probability Associated with Change in Indicated Regressor, Various Sub-Samples and Methods[1]					
	Rural			Urban		
	Probit	Probit	EA Fixed Effects	Probit	Probit	EA Fixed Effects
Number of Observations	49,554	49,554	49,554	24,539	24,539	24,539
Region of residence (excluded category is Amhara):[2]						
Tigray	0.014	0.021		0.028*	0.031*	
	(0.027)	(0.027)		(0.013)	(0.013)	
Oromiya	–0.003	–0.008		0.015	0.007	
	(0.021)	(0.020)		(0.012)	(0.012)	
Benshangul-Gumuz	0.070	0.062		0.005	0.008	
	(0.040)	(0.041)		(0.030)	(0.030)	
SNNPR	–0.012	–0.009		–0.038*	–0.054*	
	(0.020)	(0.020)		(0.014)	(0.015)	
Gambela	0.178	0.182		–0.077	–0.093	
	(0.094)	(0.099)		(0.060)	(0.061)	
Harare	–0.041	0.013		0.001	–0.013	
	(0.052)	(0.055)		(0.028)	(0.029)	
Addis Ababa	0.046	0.037		0.059*	0.036*	
	(0.133)	(0.123)		(0.015)	(0.018)	
Dire Dawa	1.000*	1.000*		0.010	0.043	
	(0.001)	(0.001)		(0.048)	(0.033)	
Number of schools in woreda per 1000 population that offer:						
Grades 1–4	0.401*	0.358*		0.104	0.063	
	(0.132)	(0.130)		(0.093)	(0.092)	
Grades 5–8	0.060	–0.023		0.043	0.016	
	(0.155)	(0.149)		(0.103)	(0.095)	
Grades 9–12	0.902	0.820		0.483	0.371	
	(0.880)	(0.964)		(0.685)	(0.655)	
Woreda density in 1000 population per sq. km.	0.436*	0.351*		–0.006*	–0.006*	
	(0.122)	(0.121)		(0.002)	(0.002)	
Woreda density squared	–31.588*	–32.353*		0.013*	0.013*	
	(15.537)	(14.984)		(0.003)	(0.003)	
Number per school in woreda of teachers for:						
Grades 1–4	–0.001	–0.003		0.002	0.001	
	(0.004)	(0.004)		(0.002)	(0.002)	
Grades 5–8	–0.001	–0.001		0.001	0.000	
	(0.004)	(0.004)		(0.002)	(0.002)	
Grades 9–12	0.002*	0.001*		0.001*	0.001*	
	(0.000)	(0.000)		(0.000)	(0.000)	
Shares of school in woreda with:						
Latrines	0.059	0.059		0.038	0.022	
	(0.044)	(0.041)		(0.028)	(0.027)	
Libraries	0.072	0.067		0.141*	0.121*	
	(0.054)	(0.054)		(0.038)	(0.038)	
Shares of woreda teachers in grades 1–8 who:						
Are female	–0.191*	–0.098		0.185*	0.181*	
	(0.084)	(0.080)		(0.055)	(0.054)	
Have less than TTI qualification	–0.393*	–0.333*		0.134	0.104	
	(0.154)	(0.143)		(0.110)	(0.095)	
Have vocational training, BA or MA	1.081	1.039		–0.132	–0.225	
	(0.992)	(0.910)		(0.289)	(0.274)	
Share of households in EA:						
With literate head		0.500*			0.113*	
		(0.045)			(0.038)	
Whose head has formal sector job		0.487			0.105*	
		(0.253)			(0.037)	

Table B.2 (continued)

	Estimated Changes in Probability Associated with Change in Indicated Regressor, Various Sub-samples and Methods[1]					
	Rural			Urban		
	Probit	OLS	EA Fixed Effects	Probit	OLS	EA Fixed Effects
Whose head has no job		0.098 (0.055)			0.069 (0.036)	
Characteristics of household head:						
Male[2]	–0.058* (0.011)	–0.048* (0.010)	–0.033* (0.007)	–0.060* (0.007)	–0.056* (0.007)	–0.059* (0.007)
Age in years	0.004* (0.001)	0.003* (0.001)	0.003* (0.001)	0.003* (0.001)	0.004* (0.001)	0.004* (0.001)
Age squared	–0.002 (0.001)	–0.002 (0.001)	–0.002* (0.001)	–0.003* (0.001)	–0.003* (0.001)	–0.004* (0.001)
Literate[2]	0.138* (0.008)	0.091* (0.007)	0.082* (0.005)	0.098* (0.008)	0.086* (0.007)	0.089* (0.006)
Reports traditional religion[2]	–0.128* (0.015)	–0.106* (0.015)	–0.061* (0.011)	–0.121* (0.060)	–0.113* (0.056)	–0.126* (0.035)
Ever migrated[2]	0.049* (0.009)	0.038* (0.009)	0.001 (0.005)	0.033* (0.008)	0.026* (0.008)	0.019* (0.006)
Has formal job[2]	0.235* (0.040)	0.177* (0.042)	0.167* (0.026)	0.061* (0.010)	0.051* (0.010)	0.048* (0.010)
Has no job[2]	–0.021 (0.012)	–0.035* (0.010)	–0.018* (0.008)	0.021* (0.007)	0.014 (0.007)	0.022* (0.007)
Household size (number of members)	0.007* (0.002)	0.006* (0.002)	0.005* (0.001)	0.008* (0.001)	0.008* (0.001)	0.009* (0.001)
Share of household members who are:						
Under 7 years old	–0.067* (0.025)	–0.048 (0.026)	–0.022 (0.018)	–0.147* (0.021)	–0.142* (0.021)	–0.129* (0.020)
Male and over 15 years	–0.116* (0.030)	–0.119* (0.030)	–0.089* (0.021)	–0.022 (0.025)	–0.026 (0.025)	–0.023 (0.022)
Female and over 15 yrs.	–0.009 (0.032)	–0.013 (0.032)	–0.014 (0.023)	0.060* (0.025)	0.054* (0.025)	0.028 (0.022)
In formal jobs	0.466* (0.133)	0.365* (0.139)	0.174 (0.097)	0.055 (0.048)	0.025 (0.047)	0.013 (0.037)
Child characteristics (excluded categories are female and age 7):[2]						
Male	0.153* (0.006)	0.155* (0.006)	0.140* (0.004)	0.047* (0.005)	0.048* (0.004)	0.050* (0.004)
Age 8	0.109* (0.010)	0.108* (0.010)	0.078* (0.007)	0.065* (0.005)	0.064* (0.005)	0.117* (0.009)
Age 9	0.210* (0.010)	0.211* (0.010)	0.154* (0.007)	0.101* (0.004)	0.101* (0.004)	0.190* (0.009)
Age 10	0.292* (0.010)	0.293* (0.010)	0.231* (0.007)	0.117* (0.004)	0.116* (0.004)	0.221* (0.008)
Age 11	0.339* (0.011)	0.340* (0.011)	0.269* (0.008)	0.121* (0.004)	0.120* (0.004)	0.243* (0.009)
Age 12	0.364* (0.010)	0.367* (0.010)	0.299* (0.007)	0.127* (0.004)	0.126* (0.004)	0.235* (0.008)
Age 13	0.390* (0.010)	0.392* (0.010)	0.319* (0.008)	0.119* (0.004)	0.118* (0.004)	0.233* (0.009)
Age 14	0.387* (0.011)	0.389* (0.011)	0.326* (0.008)	0.125* (0.004)	0.124* (0.004)	0.233* (0.009)
Whether child reports:						
Mother not living2	–0.046* (0.012)	–0.047* (0.012)	–0.040* (0.009)	–0.054* (0.014)	–0.052* (0.014)	–0.044* (0.011)
Father not living2	–0.049* (0.009)	–0.047* (0.009)	–0.049* (0.007)	–0.042* (0.009)	–0.041* (0.009)	–0.035* (0.007)
Both parents not living2	–0.019 (0.027)	–0.021 (0.027)	–0.004 (0.021)	0.001 (0.019)	–0.002 (0.019)	–0.007 (0.018)

[1,2] See Table B.1, notes 1 and 2.

* Asterisks identify coefficient estimates that are significantly different from zero at the two-tailed five-percent level.

Table B.3. Estimates of Probability Derivatives Employing the LFS/EMIS 1999
Samples: Rural and Urban Boys and Girls 7 to 14 Years Old
Dichotomous Dependent Variable: Whether Child has Ever Attended School

	Estimated Changes in Probability Associated with Change in Indicated Regressor, from Probit estimates, Various Sub-Samples[1]			
	Rural Boys	Rural Girls	Urban Boys	Urban Girls
Number of Observations	24,544	24,101	11,865	12,674
Region of residence (excluded category is Amhara):[2]				
Tigray	0.009	0.030	0.022	0.042*
	(0.031)	(0.027)	(0.014)	(0.016)
Oromiya	0.072*	–0.078*	0.014	–0.001
	(0.023)	(0.020)	(0.012)	(0.016)
Benshangul-Gumuz	0.106*	0.012	0.041	–0.035
	(0.045)	(0.043)	(0.021)	(0.044)
SNNPR	0.069*	–0.082*	–0.031*	–0.076*
	(0.024)	(0.020)	(0.015)	(0.019)
Gambela	0.119	0.224	–0.027	–0.175*
	(0.103)	(0.119)	(0.056)	(0.084)
Harare	0.140*	–0.097*	0.011	–0.038
	(0.068)	(0.042)	(0.038)	(0.040)
Addis Ababa	0.075	–0.023	0.052*	0.023
	(0.160)	(0.102)	(0.019)	(0.027)
Dire Dawa	1.000*	0.999*	0.052*	0.015
	(0.000)	(0.005)	(0.025)	(0.053)
Number of schools in woreda per 1000 pop. that offer:				
Grades 1–4	0.599*	0.157	–0.021	0.161
	(0.137)	(0.137)	(0.104)	(0.112)
Grades 5–8	–0.265	0.174	0.034	–0.024
	(0.159)	(0.157)	(0.101)	(0.122)
Grades 9–12	2.109*	–0.241	0.213	0.585
	(1.074)	(1.071)	(0.772)	(0.787)
Woreda density in 1000 population per sq. km.	0.504*	0.199	–0.005	–0.007*
	(0.137)	(0.129)	(0.003)	(0.003)
Woreda density squared	–41.918*	–22.727	0.010*	0.016*
	(16.769)	(15.859)	(0.004)	(0.004)
Number per school in woreda of teachers for:				
Grades 1–4	–0.005	0.001	0.003	0.000
	(0.005)	(0.004)	(0.002)	(0.003)
Grades 5–8	–0.007	0.005	–0.001	0.001
	(0.005)	(0.004)	(0.002)	(0.002)
Grades 9–12	0.001*	0.001*	0.000	0.001*
	(0.000)	(0.000)	(0.000)	(0.000)
Shares of school in woreda with:				
Latrines	0.064	0.047	–0.004	0.048
	(0.047)	(0.045)	(0.031)	(0.034)
Libraries	0.100	0.031	0.122*	0.117*
	(0.060)	(0.057)	(0.038)	(0.047)
Shares of woreda teachers in grades 1–8 who:				
Are female	–0.165	–0.043	0.179*	0.182*
	(0.090)	(0.084)	(0.059)	(0.067)
Have less than TTI qualification	–0.565*	–0.104	0.099	0.109
	(0.167)	(0.147)	(0.099)	(0.111)
Have vocational training, BA or MA	2.467*	–0.214	–0.035	–0.373
	(1.115)	(0.939)	(0.354)	(0.365)
Share of households in EA:				
With literate head	0.507*	0.481*	0.106*	0.109*
	(0.053)	(0.046)	(0.040)	(0.045)
Whose head has formal sector job	0.379	0.534	0.110*	0.104*
	(0.223)	(0.291)	(0.039)	(0.048)

Table B.3 (continued)

	Estimated Changes in Probability Associated with Change in Indicated Regressor, from Probit estimates, Various Sub-Samples[1]			
	Rural Boys	Rural Girls	Urban Boys	Urban Girls
Whose head has no job	0.018	0.176*	0.062	0.071
	(0.063)	(0.058)	(0.039)	(0.043)
Characteristics of household head:				
Male[2]	–0.050*	–0.044*	–0.030*	–0.084*
	(0.014)	(0.013)	(0.009)	(0.010)
Age in years	0.004*	0.002	0.001	0.007*
	(0.002)	(0.002)	(0.001)	(0.001)
Age squared	–0.003	–0.001	–0.001	–0.006*
	(0.002)	(0.002)	(0.001)	(0.001)
Literate[2]	0.085*	0.094*	0.068*	0.106*
	(0.009)	(0.009)	(0.008)	(0.010)
Reports traditional religion[2]	–0.095*	–0.120*	–0.135*	–0.079
	(0.019)	(0.015)	(0.064)	(0.062)
Ever migrated[2]	0.036*	0.041*	0.030*	0.021*
	(0.010)	(0.010)	(0.008)	(0.010)
Has formal job[2]	0.213*	0.112*	0.037*	0.062*
	(0.049)	(0.054)	(0.014)	(0.013)
Has no job[2]	–0.036*	–0.035*	0.010	0.018
	(0.014)	(0.013)	(0.009)	(0.010)
Household size (number of members)	0.006*	0.005*	0.004*	0.011*
	(0.002)	(0.002)	(0.002)	(0.002)
Share of household members who are:				
Under 7 years old	–0.004	–0.084*	–0.081*	–0.188*
	(0.035)	(0.032)	(0.025)	(0.031)
Male and over 15 years	–0.128*	–0.101*	–0.051	0.005
	(0.041)	(0.036)	(0.029)	(0.035)
Female and over 15 yrs.	–0.051	0.031	0.073*	0.037
	(0.044)	(0.039)	(0.029)	(0.035)
In formal jobs	0.160	0.605*	0.077	0.004
	(0.160)	(0.191)	(0.073)	(0.062)
Child characteristics (excluded categories are female and age 7):[2]				
Age 8	0.128*	0.082*	0.064*	0.058*
	(0.013)	(0.013)	(0.006)	(0.009)
Age 9	0.235*	0.173*	0.091*	0.104*
	(0.013)	(0.014)	(0.005)	(0.007)
Age 10	0.326*	0.240*	0.100*	0.126*
	(0.012)	(0.015)	(0.005)	(0.007)
Age 11	0.373*	0.281*	0.096*	0.141*
	(0.013)	(0.017)	(0.005)	(0.006)
Age 12	0.415*	0.292*	0.111*	0.134*
	(0.012)	(0.016)	(0.005)	(0.006)
Age 13	0.443*	0.306*	0.104*	0.125*
	(0.011)	(0.017)	(0.005)	(0.007)
Age 14	0.452*	0.279*	0.111*	0.128*
	(0.011)	(0.017)	(0.005)	(0.007)
Whether child reports:				
Mother not living[2]	–0.041*	–0.052*	–0.037*	–0.067*
	(0.017)	(0.014)	(0.017)	(0.019)
Father not living[2]	–0.059*	–0.034*	–0.046*	–0.035*
	(0.013)	(0.012)	(0.012)	(0.012)
Both parents not living[2]	–0.008	–0.029	0.003	–0.011
	(0.036)	(0.036)	(0.024)	(0.028)

[1,2] See Table B.1, notes 1 and 2. * Asterisks identify estimates that are significantly different from zero at the five-percent level.

Table B.4. Household Fixed Effects Estimates of Probability Derivatives Employing the LFS/EMIS 1999
Samples: Rural and Urban Children 7 to 14 Years Old
Dichotomous Dependent Variable: Whether Child has Ever Attended School

	Estimated Changes in Probability Associated with Change in Indicated Regressor, Probit Estimates, Rural and Urban Sub-Samples[1]	
	Rural	Urban
Number of Observations	49,554	24,539
Child characteristics (excluded categories are female and age 7):[2]		
Male	0.161*	0.043*
	(0.005)	(0.005)
Age 8	0.072*	0.126*
	(0.010)	(0.011)
Age 9	0.159*	0.212*
	(0.009)	(0.010)
Age 10	0.244*	0.249*
	(0.009)	(0.010)
Age 11	0.301*	0.271*
	(0.010)	(0.011)
Age 12	0.341*	0.269*
	(0.009)	(0.010)
Age 13	0.355*	0.260*
	(0.010)	(0.011)
Age 14	0.371*	0.260*
	(0.010)	(0.010)
Whether child reports:		
Mother not living[2]	–0.062*	–0.131*
	(0.024)	(0.023)
Father not living[2]	–0.073*	–0.114*
	(0.022)	(0.016)
Both parents not living[2]	–0.049	0.077*
	(0.048)	(0.035)

[1,2] See Table B.1, notes 1 and 2. * Asterisks identify estimates that are significantly different from zero at the five-percent level.

Table B.5. Participation Rates in Various Activities
Samples: Rural and Urban Boys and Girls 7 to 14 Years Old

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

Table B.6. Estimates of Probability Derivatives Employing the LFS/EMIS 1999
Samples: Rural and Urban Boys 7 to 14 Years Old
Dichotomous Dependent Variables Related to Activities in Last 7 Days

	Estimated Changes in Probability Associated with Indicated Regressor, from Probit Estimates, Various Sub-Samples and Dependent Variables[1]					
	Rural			Urban		
	School attendance	Income generation	House work	School attendance	Income generation	House work
Number of Observations	24,536	24,536	24,536	11,859	11,859	11,859
Number of schools in woreda per 1000 pop. that offer:						
Grades 1–4	0.447*	–0.482*	–0.078	–0.015	–0.071	0.129
	(0.125)	(0.171)	(0.126)	(0.123)	(0.133)	(0.273)
Grades 5–8	–0.198	0.807*	–0.025	–0.027	0.179	0.318
	(0.158)	(0.213)	(0.157)	(0.136)	(0.141)	(0.338)
Grades 9–12	2.187*	–2.477*	2.219*	0.391	–3.462*	–4.253*
	(0.920)	(1.223)	(0.908)	(0.906)	(1.098)	(1.747)
Woreda density in 1000	0.143*	0.026	0.001	–0.004	–0.001	–0.005
population per sq. km.	(0.049)	(0.076)	(0.055)	(0.003)	(0.003)	(0.006)
Woreda density squared	–0.307*	–0.054	–0.002	0.007	–0.000	0.010
	(0.107)	(0.167)	(0.120)	(0.004)	(0.005)	(0.009)
Number per school in woreda of teachers for:						
Grades 1–4	–0.005	–0.008	0.007	–0.003	0.001	0.002
	(0.004)	(0.006)	(0.004)	(0.003)	(0.003)	(0.006)
Grades 5–8	–0.009*	0.007	–0.009*	0.004	0.001	0.005
	(0.004)	(0.006)	(0.005)	(0.003)	(0.003)	(0.005)
Grades 9–12	0.001*	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
Shares of school in woreda with:						
Latrines	0.086*	–0.082	0.078	–0.037	0.018	0.083
	(0.044)	(0.075)	(0.053)	(0.036)	(0.044)	(0.079)
Libraries	0.064	–0.056	–0.000	0.148*	–0.027	0.009
	(0.055)	(0.080)	(0.060)	(0.045)	(0.054)	(0.103)
Shares of woreda teachers in grades 1–8 who:						
Are female	–0.127	–0.022	0.081	0.244*	–0.099	–0.213
	(0.082)	(0.124)	(0.095)	(0.070)	(0.085)	(0.155)
Have less than TTI	–0.500*	0.034	–0.103	0.138	0.121	0.263
qualification	(0.155)	(0.197)	(0.162)	(0.139)	(0.139)	(0.261)
Have vocational training,	2.152*	1.236	–1.874	0.004	–0.173	–0.298
BA or MA	(0.977)	(1.652)	(1.241)	(0.378)	(0.518)	(0.755)
Share of households in EA:						
With literate head	0.442*	–0.114	0.082	0.108*	–0.042	–0.146
	(0.049)	(0.072)	(0.053)	(0.049)	(0.048)	(0.089)
Whose head has formal	0.295	–0.731*	0.318	0.166*	–0.074	0.261*
sector job	(0.175)	(0.252)	(0.178)	(0.046)	(0.051)	(0.098)
Whose head has no job	0.028	–0.439*	0.114	0.072	–0.212*	0.089
	(0.058)	(0.082)	(0.064)	(0.046)	(0.051)	(0.103)
Characteristics of household head:						
Male[2]	–0.046*	0.017	–0.004	–0.023*	0.035*	–0.020
	(0.013)	(0.015)	(0.011)	(0.011)	(0.010)	(0.016)
Age in years	0.004*	–0.003	0.000	0.002	–0.003	0.008*
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)

Table B.6 (continued)

	Estimated Changes in Probability Associated with Indicated Regressor, from Probit Estimates, Various Sub-Samples and Dependent Variables[1]					
	Rural			Urban		
	School attendance	Income generation	House work	School attendance	Income generation	House work
Age squared	–0.003	0.004*	–0.001	–0.002	0.003	–0.007*
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)
Literate[2]	0.083*	0.000	0.019*	0.087*	–0.037*	0.051*
	(0.009)	(0.010)	(0.007)	(0.010)	(0.008)	(0.014)
Reports traditional religion[2]	–0.094*	0.021	–0.021	–0.142	–0.008	0.030
	(0.018)	(0.025)	(0.017)	(0.078)	(0.052)	(0.060)
Ever migrated[2]	0.023*	–0.016	0.014	0.024*	–0.029*	0.032*
	(0.010)	(0.012)	(0.010)	(0.011)	(0.011)	(0.016)
Has formal job[2]	0.159*	–0.118*	0.013	0.049*	–0.063*	0.075*
	(0.047)	(0.050)	(0.038)	(0.018)	(0.013)	(0.024)
Has no job[2]	–0.044*	–0.049*	0.014	0.012	–0.049*	0.008
	(0.013)	(0.015)	(0.013)	(0.011)	(0.008)	(0.016)
Household size (number of members)	0.009*	–0.012*	–0.002	0.008*	0.000	–0.006*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Share of household members who are:						
Under 7 years old	–0.037	0.317*	0.029	–0.143*	0.049	0.107*
	(0.033)	(0.036)	(0.028)	(0.034)	(0.029)	(0.049)
Male and over 15 years	–0.106*	0.120*	0.011	–0.057	–0.002	0.102*
	(0.039)	(0.041)	(0.034)	(0.039)	(0.031)	(0.052)
Female and over 15 yrs.	–0.063	0.168*	–0.085*	0.110*	–0.101*	0.010
	(0.042)	(0.045)	(0.035)	(0.037)	(0.033)	(0.051)
In formal jobs	0.230	0.028	0.080	0.134	–0.035	–0.107
	(0.146)	(0.175)	(0.143)	(0.086)	(0.070)	(0.090)
Child characteristics (excluded categories are female and age 7):[2]						
Age 8	0.118*	0.094*	0.038*	0.087*	0.005	0.057*
	(0.013)	(0.012)	(0.010)	(0.008)	(0.018)	(0.018)
Age 9	0.230*	0.176*	0.059*	0.121*	0.081*	0.141*
	(0.013)	(0.012)	(0.011)	(0.007)	(0.021)	(0.019)
Age 10	0.307*	0.342*	–0.037*	0.133*	0.188*	0.136*
	(0.013)	(0.011)	(0.011)	(0.006)	(0.027)	(0.020)
Age 11	0.355*	0.364*	–0.047*	0.131*	0.250*	0.151*
	(0.014)	(0.011)	(0.011)	(0.007)	(0.030)	(0.022)
Age 12	0.385*	0.393*	–0.054*	0.144*	0.295*	0.185*
	(0.013)	(0.011)	(0.010)	(0.007)	(0.027)	(0.020)
Age 13	0.402*	0.411*	–0.063*	0.123*	0.315*	0.180*
	(0.013)	(0.010)	(0.011)	(0.007)	(0.030)	(0.021)
Age 14	0.407*	0.427*	–0.085*	0.129*	0.351*	0.165*
	(0.013)	(0.010)	(0.010)	(0.007)	(0.029)	(0.021)
Whether child reports:						
Mother not living[2]	–0.055*	0.038*	–0.022	–0.061*	0.004	–0.011
	(0.015)	(0.017)	(0.013)	(0.020)	(0.015)	(0.024)
Father not living[2]	–0.060*	0.032*	–0.006	–0.063*	0.030*	0.012
	(0.012)	(0.014)	(0.011)	(0.014)	(0.011)	(0.016)
Both parents not living[2]	–0.041	0.002	0.004	0.015	–0.034	0.006
	(0.035)	(0.038)	(0.030)	(0.028)	(0.020)	(0.042)

[1,2] See Table B.1, notes 1 and 2. Probit regressions also include indicators of region of residence.

Table B.7. Estimates of Probability Derivatives Employing the LFS/EMIS 1999
Samples: Rural and Urban Girls 7 to 14 Years Old
Dichotomous Dependent Variables Related to Activities in Last 7 Days

	Estimated Changes in Probability Associated with Indicated Regressor, from Probit Estimates, Various Sub-Samples and Dependent Variables[1]					
	Rural			Urban		
	School attendance	Income generation	House work	School attendance	Income generation	House work
Number of Observations	24,002	24,002	24,002	12,665	12,665	12,665
Number of schools in woreda per 1000 pop. that offer:						
Grades 1–4	0.085	0.038	–0.210	0.132	–0.167	–0.045
	(0.119)	(0.142)	(0.155)	(0.131)	(0.087)	(0.288)
Grades 5–8	0.188	0.028	0.259	–0.187	0.227*	0.327
	(0.139)	(0.185)	(0.183)	(0.181)	(0.096)	(0.367)
Grades 9–12	–0.660	–0.008	0.946	0.402	–1.216*	–2.803
	(0.945)	(1.084)	(1.110)	(0.934)	(0.557)	(1.786)
Woreda density in 1000	0.001	–0.005	0.037	–0.007*	0.003	–0.006
population per sq. km.	(0.046)	(0.062)	(0.067)	(0.003)	(0.002)	(0.007)
Woreda density squared	0.003	0.009	–0.076	0.015*	–0.006	0.014
	(0.101)	(0.135)	(0.146)	(0.004)	(0.003)	(0.013)
Number per school in woreda of teachers for:						
Grades 1–4	–0.000	–0.004	–0.001	–0.003	–0.000	0.006
	(0.004)	(0.006)	(0.005)	(0.003)	(0.002)	(0.006)
Grades 5–8	0.004	0.009	–0.010*	0.001	0.001	0.002
	(0.004)	(0.005)	(0.005)	(0.003)	(0.002)	(0.005)
Grades 9–12	0.001*	–0.000	0.001	0.001*	–0.000	0.000
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)
Shares of school in woreda with:						
Latrines	0.071	–0.119*	0.162*	0.010	0.018	0.042
	(0.044)	(0.059)	(0.058)	(0.039)	(0.024)	(0.086)
Libraries	0.025	0.016	–0.063	0.185*	–0.003	–0.076
	(0.055)	(0.069)	(0.069)	(0.050)	(0.031)	(0.109)
Shares of woreda teachers in grades 1–8 who:						
Are female	–0.031	–0.163	0.119	0.223*	–0.042	–0.140
	(0.078)	(0.101)	(0.105)	(0.076)	(0.050)	(0.162)
Have less than TTI	–0.170	0.208	–0.302	0.119	0.114	0.036
qualification	(0.136)	(0.176)	(0.199)	(0.120)	(0.067)	(0.291)
Have vocational training,	–0.041	–1.091	0.378	0.020	0.280	–0.596
BA or MA	(0.837)	(1.466)	(1.515)	(0.376)	(0.209)	(0.848)
Share of households in EA:						
With literate head	0.443*	–0.150*	0.123*	0.134*	–0.004	–0.068
	(0.044)	(0.059)	(0.062)	(0.048)	(0.028)	(0.093)
Whose head has formal	0.360	–0.240	0.069	0.150*	–0.077*	0.137
sector job	(0.237)	(0.250)	(0.222)	(0.052)	(0.034)	(0.095)
Whose head has no job	0.169*	–0.546*	0.157*	0.089	–0.149*	0.144
	(0.052)	(0.080)	(0.074)	(0.050)	(0.031)	(0.108)
Characteristics of household head:						
Male[2]	–0.035*	0.016	–0.017	–0.091*	0.000	0.004
	(0.012)	(0.012)	(0.013)	(0.012)	(0.007)	(0.017)
Age in years	0.002	–0.001	–0.001	0.007*	–0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Age squared	–0.001	0.002	0.002	–0.006*	0.002	–0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)

Table B.7 (continued)

	Estimated Changes in Probability Associated with Indicated Regressor, from Probit Estimates, Various Sub-Samples and Dependent Variables[1]					
	Rural			Urban		
	School attendance	Income generation	House work	School attendance	Income generation	House work
Literate[2]	0.086*	–0.013	0.025*	0.126*	–0.020*	0.030*
	(0.009)	(0.009)	(0.009)	(0.011)	(0.007)	(0.013)
Reports traditional religion[2]	–0.102*	0.027	–0.036	–0.089	–0.026	0.023
	(0.014)	(0.022)	(0.022)	(0.080)	(0.034)	(0.081)
Ever migrated[2]	0.030*	–0.021*	0.019	0.020	0.004	0.014
	(0.009)	(0.010)	(0.011)	(0.011)	(0.007)	(0.016)
Has formal job[2]	0.114*	–0.053	–0.026	0.072*	–0.035*	0.032
	(0.049)	(0.047)	(0.055)	(0.016)	(0.009)	(0.023)
Has no job[2]	–0.039*	–0.041*	0.015	0.012	–0.037*	0.045*
	(0.012)	(0.013)	(0.014)	(0.012)	(0.005)	(0.015)
Household size (number of members)	0.008*	–0.005*	–0.011*	0.014*	–0.003*	–0.011*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.003)
Share of household members who are:						
Under 7 years old	–0.103*	0.148*	0.172*	–0.245*	0.039	0.172*
	(0.030)	(0.032)	(0.035)	(0.035)	(0.022)	(0.048)
Male and over 15 years	–0.105*	0.048	0.050	0.017	–0.049*	0.087
	(0.033)	(0.035)	(0.039)	(0.040)	(0.024)	(0.053)
Female and over 15 yrs.	0.061	0.242*	–0.234*	0.043	–0.060*	–0.115*
	(0.037)	(0.041)	(0.045)	(0.040)	(0.022)	(0.050)
In formal jobs	0.475*	–0.251	0.448*	0.028	0.069	–0.049
	(0.161)	(0.201)	(0.222)	(0.069)	(0.040)	(0.082)
Child characteristics (excluded categories are female and age 7):[2]						
Age 8	0.071*	0.102*	0.074*	0.066*	0.026	0.100*
	(0.013)	(0.014)	(0.012)	(0.011)	(0.016)	(0.018)
Age 9	0.158*	0.149*	0.166*	0.118*	0.041*	0.212*
	(0.014)	(0.014)	(0.013)	(0.009)	(0.017)	(0.017)
Age 10	0.211*	0.359*	0.031*	0.141*	0.147*	0.230*
	(0.015)	(0.016)	(0.015)	(0.009)	(0.022)	(0.017)
Age 11	0.254*	0.385*	0.052*	0.162*	0.188*	0.253*
	(0.017)	(0.017)	(0.017)	(0.008)	(0.027)	(0.017)
Age 12	0.255*	0.425*	0.049*	0.144*	0.263*	0.281*
	(0.016)	(0.016)	(0.016)	(0.009)	(0.026)	(0.016)
Age 13	0.255*	0.467*	0.046*	0.130*	0.323*	0.263*
	(0.017)	(0.016)	(0.017)	(0.009)	(0.028)	(0.017)
Age 14	0.225*	0.509*	0.007	0.107*	0.360*	0.259*
	(0.017)	(0.015)	(0.018)	(0.010)	(0.029)	(0.017)
Whether child reports:						
Mother not living2	–0.057*	–0.005	0.028	–0.080*	–0.004	0.009
	(0.013)	(0.016)	(0.016)	(0.021)	(0.011)	(0.025)
Father not living2	–0.037*	0.006	0.007	–0.044*	0.014*	0.016
	(0.011)	(0.012)	(0.013)	(0.013)	(0.007)	(0.015)
Both parents not living2	–0.020	–0.024	–0.017	–0.025	–0.002	–0.009
	(0.033)	(0.034)	(0.038)	(0.033)	(0.018)	(0.040)

[1,2] See Table B.1, notes 1 and 2. Probit regressions also include indicators of region of residence.

* Asterisks identify estimates that are significantly different from zero at the five-percent level.

Table B.8. Percentage of Distribution of Responses Regarding Reasons for Migrating from LFS 1999
Sample: Individuals Who Have Migrated in Last Five Years

	Rural		Urban	
Reasons for migrating	Children 7–14 years old	Heads of households of children 7 to 14 years old	Children 7–14 years old	Heads of households of children 7 to 14 years old
Education	3.8	1.6	24.4	7.4
Marriage arrangement	2.2	6.9	0.1	2.5
Marriage dissolution	0.6	3.4	0.2	3.8
Search for work	13.1	9.9	10.0	18.0
Job transfer/got job	0.6	7.0	0.5	27.4
War/drought	3.8	7.3	3.2	7.2
Along with family	48.6	3.1	43.4	3.0
Return back to home	8.8	27.8	6.2	10.8
Shortage of land	0.0	9.5	0.0	1.6
To live with relatives	14.1	15.8	8.2	10.3
Health problems	0.0	0.5	0.1	2.2
Lost family	0.8	2.7	1.1	0.1
Other	1.9	2.7	2.2	5.3
Non-response	1.7	1.8	0.2	0.6

Analysis of the DHS 2000 Data

This appendix presents some basic results regarding the determinants of whether children have ever attended school, for the DHS 2000 data. Because many of the most interesting variables are available only for children whose mothers responded to the female questionnaire, I restrict attention to that sub-sample. Only women between 15 and 49 years of age were interviewed. Thus we are restricting attention to children whose mothers are alive, living in the household, and in this age range.

The dependent variable employed below is an indicator of whether or not the child has ever attended school, the construction of which is discussed in the main text. The first two columns of Table C.1 present descriptive statistics, within rural and urban sub-samples, for the regressors employed in the analysis.

Region Indicators. The region variables are indicators pertaining to the administrative region of residence of the child. The excluded region is Amhara. Because the DHS contains no explicit school supply variables, these regional indicators are likely to pick up average differences across regions in school registration rates arising out of differences in school supply characteristics. Unfortunately, they also pick up the effects of systematic differences across regions in any other economic and social circumstances affecting child schooling, thus it is difficult to interpret the coefficients on the regional variables.

Asset variables. The asset variables are simple indicators of the level and types of assets owned by the household. The DHS contains no explicit measure of level of living, such as consumption expenditure per capita. The asset variables together proxy for level of living, though their coefficient estimates cannot, unfortunately, be used to calculate an explicit estimate of the effect of increased income or consumption expenditure per capita on school registration. In addition to proxying for the level of consumption expenditure per capita, the asset variables may also pick up the effects of variation in opportunity costs of child time. Some assets may increase the value of a child's time in work in a household enterprise, tending to reduce the probability of schooling, holding all else (including the level of living) constant. The first four assets are much more prevalent in urban areas, while the remaining assets are more prevalent in rural areas. The asset variables are thus highly collinear with urbanicity, and tend to capture not just variation in household resource levels, but also variation in the extent to which the household lives in an urban setting.

Household size and structure variables. After controlling at least implicitly for consumption expenditure per capita, schooling rates might rise with household size, if there are economies of scale in consumption. Household structure (e.g. the fraction of household members who are young children, and who are men and women over 14 years of age) may influence the opportunity cost of a school-age child's time. More young children means more child care needs within the household. More adults may mean greater potential to replace school-aged children by adults as child-care givers.

Language. The language of the child's mother—which we take to be the primary language of the child—may influence the probability of schooling, if schools are more attractive to those whose languages are taken into account by school systems. Until recently schools have been conducted primarily in Amharic (Amarigna) and English. The excluded language category includes many other, more minor languages.

Parental variables. The parental variables pick up a variety of characteristics of the child's mother and father. Attention is restricted to children whose mothers are present in the household (as discussed above), but some have no father. Mother's age (and its square) may pick up changes in schooling rates as they evolve over a household's life cycle, as well as cohort effects.

The schooling levels of a child's mother and father are thought to pick up changes in both attitudes and income (if income is not perfectly controlled for by, say, consumption expenditure) associated with obtaining schooling. They are often found to have strong effects on child schooling. Because Ethiopian parents tend to have very little education, and many have none at all, these variables are simply indicators of whether or not the parents have completed at least first grade.

Additional characteristics of the mother are: whether she was born in a rural area, whether her husband has other wives (set to 0 for the unmarried). Among polygamous households, an indicator captures whether the wife's rank is lower than first (set to 0 for the unmarried or married women in monogamous households).

Child variables. Among the child variables are indicators of the child's sex and age, as well as indicators of whether the child is the oldest or youngest child of his/her mother, whether among all children or among only male children. These are meant to capture any norms regarding the allocation of schooling and other rights and responsibilities across siblings. Being the youngest may also reflect the effect of having older siblings who can help support the child's schooling.

Attitude variables. The attitude variables are more unconventional than the preceding variables for inclusion in a study like this. They are included in only some specifications, because one might fear that they are endogenous, in the sense that they are jointly determined with child schooling choices. Their coefficients should not be taken as measures of causal effects on school registration probabilities. Rather, their coefficients indicate whether there are important co-movements between these "attitude" variables and child schooling, after controlling for other factors in the table. The variables are: whether or not the mother listens to the radio at least once a week, whether the mother responded that it is justified for a husband to beat his wife if she goes out without his permission, and whether the mother reports her religion to be something other than Orthodox, Catholic, Protestant or Muslim. (This last one is less likely to be endogenous than the previous two.)

Basic Multivariate Results. Tables C.1 and C.2 report estimates of coefficients in a variety of specifications relating the probability that a child has ever attended school to some or all of the potential determinants. All specifications in Table C.1 employ the entire sample, while the specifications in Table C.2 pertain to rural and urban sub-samples.

The main observations from Table C.1 are the following:

- The overall difference in school registration probabilities between rural and urban areas is very large, and does not diminish much when regional controls are added. Thus the overall rural–urban difference is not just picking up a tendency for registration rates to differ between regions that are more and less urban. The rural–urban difference is profound within regions.

- The addition of household, parent and child-level controls, however, diminishes the rural–urban difference quite a bit. Other regressions not shown suggest that the household asset variables (which proxy for consumption expenditure per capita as well as opportunity costs) play the largest role in reducing the residual rural–urban difference. In light of the results derived with the WMS/HICES data, which include direct controls for household resource levels, it seems likely that the results here overstate the role of household wealth in explaining rural–urban differences in school attendance rates, because the asset variables proxy not only for resources but also for urbanicity itself.

- The smaller magnitude of some estimated asset variable effects in the enumeration area fixed effects specification (which essentially introduces a dummy variable for every enumeration area in the dataset, and should thus control much better for degree of urbanicity of the location in which a household resides), is consistent with the view that they capture geographic differences in addition to differences in household resources.

- The fact that many household asset variables retain significance even in the household fixed effects specification, however, suggests that they also capture household-level variation in resource levels, and that these resource levels matter to school attendance.

- Consistent with the results in other datasets, children from households in which there are more adult males are less likely to attend school, while children in households with more adult females are more likely to attend school. This is consistent with the possibility that the presence of more adult males in the household increases the productivity of family enterprises, increasing the opportunity cost of the child's time, while increases in the numbers of adult females frees children up from housework responsibilities to attend school. The variable describing the share of household members who are under age 7 does not have a significant effect in these specifications, in contrast to the negative effect found in other datasets. Note, however, that this specification also controls for whether the child to which the observation pertains is the youngest among his or her siblings. This picks up some of the same effect.

- Children speaking Amarigna and Tigrigna are more likely to register for school than children speaking other languages.

- In specifications that do not allow for enumeration area fixed effects, parental schooling plays its typical large role in explaining child schooling, suggesting a great tendency for inequality to be transmitted across generations. Note, however, that the parental education effects are greatly reduced when enumeration area fixed effects are included. This suggests that much of the standard association between parental schooling and child schooling is picking up the effect of living in communities where adults are more and less educated. This raises the possibility that what matters in not so much whether a child's own parents are educated, but whether he has exposure to educated adults in his community. This raises some hope that the probability of children's schooling can be increased through information campaigns that increase families' exposure to education and its potential benefits.

- Male children are more likely to be registered, especially in rural areas, though the differences are not terribly large by world standards.

- The probability of registration rises with age until about age 13, reflecting the late entry of many Ethiopian children into the schooling system.

- The only birth order variable of some significance is whether the child is the youngest. This may capture not only cultural tendencies to "spoil the baby," but also the beneficial effect of having older siblings who can help support a child's schooling, and of having no younger siblings who require child care.

- Children of mothers that listen to the radio are noticeably more likely to attend school than other children. While this does not necessarily mean that expanding the use of radios would increase schooling, it does suggest that the same attitudes and beliefs that cause households to want to be connected to the rest of the world via radio also tend to be associated with school registration.

- In rural areas, households that are more traditional, as indicated by the wife's belief that beating is justified if she goes out without her husband's permission, are less likely to send their children to school. This holds even in the fixed effects specification, suggesting that this is not just picking up differences across more and less modern communities. It is picking up differences across households with different attitudes even within communities. It thus supports the notion that attitudes (quite broadly defined) do matter in schooling decisions.

The main observations from Table C.3 are:

- The effects of several asset variables on school attendance rates are stronger in rural than in urban areas. Whether this reflects a stronger household resource effect, or a stronger role for the assets to play in describing variation in urbanicity within sub-samples, is unclear.

- The negative effect of the presence of more adult males on school attendance rates is primarily a rural phenomenon.

- The positive effect of the presence of more adult females on school attendance rates is also stronger in rural than urban areas.

- The effect of not having a father is stronger in urban areas, and the magnitude of the effect is similar to what was found in the LFS data.

- The effects of parental schooling on child school attendance are stronger in rural than urban areas.

- Gender differences are stronger in rural than urban areas.

- Being the youngest among siblings has a much stronger positive effect on school attendance in rural than urban areas.

- The association of school attendance rates with whether or not the mother listens to the radio is equally strong in urban and rural areas, despite large differences in the incidence of radio listening.

Subsidiary Observations

The DHS provides some information on the work activities of the mothers. The following statistics pertain to women ages 15 to 49 who responded to the female questionnaire and who are mothers of children aged 7–14.

- 65.6 (64.3) percent in rural (urban) areas report that they either are working now or have worked in the last 12 months

- Among those who work, 51.5 (51.6) percent report that they are self-employed in rural (urban) areas. Most of the rural women who work for others say they are working for family members, while most of the urban women who are working for others are not working for family members.

- While over 51 percent of women in both rural and urban areas report being self-employed, 86 (63) percent in rural (urban) areas report working "away from home." Thus even in their self-employment and farm activities, they have to work at some distance from the home. This creates a potential need for school aged children to care for younger siblings, a use of a child's time that may competes with schooling.

Table C.1. Descriptive Statistics and Estimates of Probability Derivatives Employing the DHS 2000 Sample: Children 7 to 14 Years Old whose Mothers Responded to Female Questionnaire Dichotomous Dependent Variable: Whether Child Ever Attended School

			Estimated Changes in Probability Associated with Change in Indicated Regressor, Various Methods[1]				
	Rural Means (Std. Dev.)	Urban Means (Std. Dev.)	Probit	Probit	Probit	OLS	EA Fixed Effects
Number of Observations	8,390	1,828	10,215	10,215	10,215	10,215	10,215
Urban residence[2]	0.273	0.845	0.514*	0.199*	0.185*	0.365*	
			(0.024)	(0.049)	(0.050)	(0.035)	
Region of residence (excluded category is Amhara)							
Tigray	0.062	0.131	–0.062	–0.099	–0.097	–0.083	
			(0.040)	(0.069)	(0.069)	(0.067)	
Affar	0.010	0.007	–0.137*	–0.057	–0.066	–0.041	
			(0.049)	(0.064)	(0.064)	(0.063)	
Oromiya	0.385	0.282	–0.052	–0.058	–0.063	–0.015	
			(0.034)	(0.047)	(0.046)	(0.046)	
Somali	0.013	0.037	–0.343*	–0.287*	–0.299*	–0.298*	
			(0.031)	(0.042)	(0.040)	(0.039)	
Benshangul-Gumuz	0.011	0.007	0.032	0.097	0.091	0.117*	
			(0.049)	(0.054)	(0.054)	(0.053)	
SNNPR	0.232	0.122	–0.007	0.043	0.030	0.073	
			(0.038)	(0.055)	(0.055)	(0.055)	
Gambela	0.002	0.004	0.206*	0.257*	0.262*	0.294*	
			(0.052)	(0.066)	(0.069)	(0.065)	
Harare	0.001	0.007	0.187*	0.181*	0.161*	0.256*	
			(0.041)	(0.058)	(0.058)	(0.053)	
Addis Ababa	0.000	0.198	0.219*	–0.028	–0.037	0.149*	
			(0.059)	(0.069)	(0.068)	(0.069)	
Dire Dawa	0.002	0.019	–0.077	–0.123*	–0.128*	–0.035	
			(0.045)	(0.062)	(0.061)	(0.057)	
Household Assets							
Number of sleeping rooms	1.299	1.625		0.059*	0.056*		0.032*
	(0.522)	(0.723)		(0.015)	(0.014)		(0.008)
Floor not of earth or dung[2]	0.022	0.323		0.193*	0.177*		0.058*
				(0.051)	(0.051)		(0.022)
Has electricity[2]	0.004	0.763		0.265*	0.259*		0.213*
				(0.036)	(0.036)		(0.033)
Has toilet[2]	0.080	0.711		0.094*	0.084*		0.061*
				(0.028)	(0.028)		(0.016)
Has land[2]	0.959	0.183		–0.022	–0.016		–0.011
				(0.033)	(0.033)		(0.018)
Has animals[2]	0.892	0.402		0.024	0.019		0.047*
				(0.022)	(0.022)		(0.013)
Has cash crop[2]	0.333	0.034		0.012	0.005		0.014
				(0.023)	(0.023)		(0.013)
Number of household members	7.139	6.850		0.001	0.000	0.008	0.000
	(2.141)	(2.459)		(0.004)	(0.004)	(0.004)	(0.002)
Share of household members who are:							
Under 7 years old	0.222	0.152		0.017	0.020	–0.031	0.024
	(.133)	(0.133)		(0.071)	(0.071)	(0.070)	(0.043)
Male and over 15 years	0.218	0.209		–0.068	–0.081	–0.076	–0.114*
	(.117)	(0.131)		(0.076)	(0.076)	(0.075)	(0.044)
Female and over 15 yrs.	0.218	0.284		0.233*	0.236*	0.266*	0.033
	(.097)	(0.124)		(0.083)	(0.083)	(0.081)	(0.050)
Language (excluded are languages other than those listed here):[2]							
Amarigna	0.302	0.526		0.190*	0.175*	0.193*	0.088*
				(0.042)	(0.043)	(0.042)	(0.022)
Oromigna	0.357	0.207		0.084	0.074	0.047	0.017
				(0.043)	(0.043)	(0.041)	(0.023)
Tigrigna	0.062	0.142		0.176*	0.143	0.164*	0.108*
				(0.075)	(0.074)	(0.072)	(0.041)
Parental Variables							
Has no father[2]	0.081	0.139		–0.024	–0.025	–0.029	–0.030*
				(0.023)	(0.023)	(0.022)	(0.014)

Table C.1 (continued)

	Rural Means (Std. Dev.)	Urban Means (Std. Dev.)	Estimated Changes in Probability Associated with Change in Indicated Regressor, Various Methods[1] Probit	Probit	Probit	OLS	EA Fixed Effects
Mother's age in years	36.653	35.974		0.028	0.029*	0.032	0.006
	(6.638)	(6.017)		(0.014)	(0.015)	(0.017)	(0.007)
Mother's age squared				0.000*	0.000*	0.000	0.000
				(0.000)	(0.000)	(0.000)	(0.000)
Father completed grade 1[2]	0.177	0.456		0.184*	0.176*	0.194*	0.076*
				(0.021)	(0.021)	(0.021)	(0.011)
Mother completed grade 1[2]	0.075	0.463		0.138*	0.117*	0.142*	0.025
				(0.028)	(0.028)	(0.027)	(0.016)
Mother born in rural area[2]	0.980	0.546		–0.097*	–0.087*	–0.134*	–0.024
				(0.036)	(0.036)	(0.038)	(0.017)
Husband of mother has more than one wife[2]	0.159	0.065		0.016	0.012	0.009	0.014
				(0.026)	(0.026)	(0.026)	(0.016)
Mother is wife of rank second or higher[2]	0.076	0.035		–0.016	–0.016	–0.015	–0.039
				(0.031)	(0.032)	(0.032)	(0.021)
Child characteristics (excluded categories are female and age 7):[2]							
Male	0.517	0.490		0.096*	0.095*	0.093*	0.073*
				(0.013)	(0.013)	(0.013)	(0.008)
Age 8	0.153	0.166		0.178*	0.179*	0.168*	0.090*
				(0.025)	(0.025)	(0.025)	(0.014)
Age 9	0.144	0.111		0.296*	0.299*	0.290*	0.161*
				(0.023)	(0.024)	(0.024)	(0.014)
Age 10	0.126	0.141		0.376*	0.377*	0.365*	0.233*
				(0.023)	(0.023)	(0.023)	(0.014)
Age 11	0.098	0.118		0.442*	0.442*	0.432*	0.286*
				(0.021)	(0.022)	(0.022)	(0.016)
Age 12	0.118	0.123		0.481*	0.481*	0.470*	0.310*
				(0.020)	(0.021)	(0.021)	(0.015)
Age 13	0.109	0.116		0.481*	0.482*	0.474*	0.323*
				(0.021)	(0.021)	(0.021)	(0.016)
Age 14	0.091	0.109		0.487*	0.486*	0.480*	0.329*
				(0.023)	(0.023)	(0.024)	(0.017)
Whether among children of mother this child is (excluded categories are female and neither oldest nor youngest):[2]							
Oldest	0.209	0.235		0.010	0.012	0.017	0.009
				(0.019)	(0.019)	(0.020)	(0.011)
Oldest male	0.337	0.352		0.017	0.018	0.014	0.008
				(0.013)	(0.013)	(0.013)	(0.008)
Youngest	0.105	0.211		0.089*	0.087*	0.090*	0.060*
				(0.022)	(0.022)	(0.022)	(0.014)
Youngest male	0.338	0.382		0.011	0.011	0.011	0.004
				(0.013)	(0.013)	(0.013)	(0.008)
Attitude Variables							
Whether mother:							
Listens to radio[2]	0.188	0.707			0.095*	0.118*	0.054*
					(0.017)	(0.017)	(0.011)
Believes beating is justified for going out without husband's permission2	0.629	0.407			–0.024	–0.029*	–0.022*
					(0.015)	(0.014)	(0.009)
Reports religion other than Orthodox, Catholic, Protestant or Muslim2	0.034	0.000			–0.074	–0.075	–0.029
					(0.043)	(0.042)	(0.030)

[1] For continuous regressors, the estimates are of the derivative of the probability of the dichotomous dependent variable equalling one with respect to the regressor, evaluated at the means of all right hand side variables. For dichotomous regressors, the estimates are of the change in probability as the variable is changed from zero to one, while holding all other right hand side variables at their means. Robust standard errors are shown in parentheses.ses.

[2] These regressors are dichotomous, taking the value 1 if the indicated condition is true, and zero otherwise. They capture differences in the dependent variable between the indicated category and the excluded category. Where no excluded category is explicitly mentioned in the table, it is the opposite of the category indicated.

* Asterisks identify coefficient estimates that are significantly different from zero at the two-tailed five percent level.

Table C.2. Estimates of Probability Derivatives Employing the DHS 2000
Sample: Children 7 to 14 Years Old whose Mothers Responded to Female Questionnaire
Dichotomous Dependent Variable: Whether Child Ever Attended School

	Estimated Changes in Probability Associated with Indicated Regressor, Various Sub-Samples and Methods[1]					
	Rural			Urban		
	Probit	Probit	EA Fixed Effects	Probit	Probit	EA Fixed Effects
Number of Observations	8,387	8,387	8,387	1,828	1,828	1,828
Household Assets						
Number of sleeping rooms	0.044*	0.043*	0.035*	0.033*	0.030*	0.016
	(0.013)	(0.013)	(0.010)	(0.012)	(0.012)	(0.013)
Floor not of earth or dung[2]	0.370*	0.351*	0.189*	0.027	0.020	0.019
	(0.072)	(0.073)	(0.043)	(0.015)	(0.015)	(0.022)
Has electricity[2]	0.301*	0.304*	0.166*	0.138*	0.122*	0.250*
	(0.078)	(0.078)	(0.063)	(0.037)	(0.035)	(0.032)
Has toilet[2]	0.096*	0.087*	0.069*	0.030	0.025	0.052*
	(0.030)	(0.030)	(0.021)	(0.018)	(0.017)	(0.022)
Has land[2]	–0.031	–0.025	–0.009	–0.015	–0.013	–0.043
	(0.035)	(0.035)	(0.023)	(0.020)	(0.018)	(0.027)
Has animals[2]	0.016	0.013	0.052*	0.010	0.006	0.021
	(0.019)	(0.020)	(0.015)	(0.015)	(0.014)	(0.023)
Has cash crop[2]	0.001	–0.003	0.010	0.049*	0.045*	0.130
	(0.019)	(0.019)	(0.014)	(0.011)	(0.011)	(0.069)
Number of household members	0.003	0.002	0.000	–0.004	–0.004	0.001
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
Share of household members who are:						
Under 7 years old	–0.030	–0.030	0.017	0.050	0.054	0.032
	(0.060)	(0.060)	(0.049)	(0.067)	(0.065)	(0.081)
Male and over 15 years	–0.095	–0.110	–0.121*	0.058	0.074	–0.015
	(0.064)	(0.064)	(0.051)	(0.061)	(0.059)	(0.075)
Female and over 15 yrs.	0.160*	0.161*	0.044	0.060	0.070	0.037
	(0.071)	(0.071)	(0.060)	(0.071)	(0.066)	(0.082)
Language (excluded are languages other than those listed here):[2]						
Amarigna	0.154*	0.139*	0.130*	0.063*	0.058*	0.038
	(0.048)	(0.047)	(0.032)	(0.020)	(0.020)	(0.027)
Oromigna	0.067	0.060	0.016	–0.002	–0.005	–0.016
	(0.048)	(0.047)	(0.029)	(0.015)	(0.015)	(0.031)
Tigrigna	0.038	0.019	0.114	0.058*	0.051*	0.065
	(0.078)	(0.074)	(0.074)	(0.011)	(0.011)	(0.042)
Parental Variables						
Has no father[2]	–0.004	–0.006	–0.026	–0.043	–0.040	–0.052*
	(0.021)	(0.021)	(0.017)	(0.023)	(0.022)	(0.024)
Mother's age in years	0.022	0.023	0.006	0.011	0.011	0.024
	(0.014)	(0.015)	(0.008)	(0.009)	(0.008)	(0.014)
Mother's age 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)
Father completed grade 1[2]	0.173*	0.165*	0.086*	0.040*	0.035*	0.035
	(0.022)	(0.022)	(0.014)	(0.013)	(0.013)	(0.018)
Mother completed grade 1[2]	0.129*	0.113*	0.045*	0.051*	0.037*	0.016
	(0.031)	(0.031)	(0.021)	(0.017)	(0.015)	(0.020)
Mother born in rural area[2]	–0.181*	–0.167*	–0.092*	–0.012	–0.013	–0.016
	(0.052)	(0.052)	(0.030)	(0.013)	(0.012)	(0.017)
Husband of mother has more than one wife[2]	0.010	0.007	0.010	0.020	0.016	0.083
	(0.022)	(0.022)	(0.017)	(0.022)	(0.024)	(0.058)
Mother is wife of rank second or higher[2]	–0.020	–0.019	–0.035	0.024	0.026	–0.058
	(0.026)	(0.026)	(0.022)	(0.022)	(0.020)	(0.072)

Table B.2 (continued)

	Estimated Changes in Probability Associated with Change in Indicated Regressor, Various Sub-samples and Methods[1]					
	Rural			Urban		
	Probit	OLS	EA Fixed Effects	Probit	OLS	EA Fixed Effects
Child characteristics (excluded categories are female and age 7):[2]						
Male	0.085*	0.084*	0.084*	0.012	0.012	0.022
	(0.012)	(0.012)	(0.009)	(0.011)	(0.011)	(0.015)
Age 8	0.139*	0.138*	0.077*	0.049*	0.047*	0.166*
	(0.025)	(0.025)	(0.015)	(0.010)	(0.009)	(0.028)
Age 9	0.251*	0.254*	0.153*	0.061*	0.058*	0.225*
	(0.026)	(0.026)	(0.016)	(0.009)	(0.009)	(0.030)
Age 10	0.332*	0.332*	0.225*	0.078*	0.074*	0.276*
	(0.028)	(0.028)	(0.016)	(0.010)	(0.010)	(0.028)
Age 11	0.422*	0.422*	0.289*	0.073*	0.069*	0.287*
	(0.028)	(0.028)	(0.018)	(0.010)	(0.010)	(0.030)
Age 12	0.466*	0.467*	0.320*	0.080*	0.076*	0.275*
	(0.027)	(0.027)	(0.017)	(0.011)	(0.011)	(0.029)
Age 13	0.467*	0.468*	0.334*	0.079*	0.075*	0.285*
	(0.028)	(0.028)	(0.018)	(0.011)	(0.010)	(0.031)
Age 14	0.480*	0.480*	0.347*	0.075*	0.070*	0.272*
	(0.031)	(0.032)	(0.019)	(0.010)	(0.010)	(0.032)
Whether among children of mother this child is (excluded categories are female and neither oldest nor youngest):[2]						
Oldest	0.008	0.009	0.005	0.002	0.002	0.027
	(0.017)	(0.017)	(0.013)	(0.014)	(0.013)	(0.020)
Oldest male	0.016	0.016	0.010	0.000	0.000	0.005
	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)	(0.016)
Youngest	0.093*	0.091*	0.074*	0.010	0.008	0.006
	(0.022)	(0.022)	(0.017)	(0.013)	(0.013)	(0.021)
Youngest male	0.006	0.006	0.007	0.008	0.008	0.005
	(0.012)	(0.012)	(0.009)	(0.010)	(0.010)	(0.015)
Attitude Variables						
Whether mother:						
Listens to radio[2]		0.068*	0.050*		0.049*	0.068*
		(0.016)	(0.012)		(0.017)	(0.020)
Believes beating is justified for going out without husband's permission[2]		–0.025 (0.013)	–0.027* (0.010)		–0.003 (0.012)	–0.001 (0.018)
Reports religion other than Orthodox, Catholic, Protestant or Muslim[2]		–0.056 (0.033)	–0.022 (0.032)			–0.333 (0.222)

[1,2] See Table C.1, notes 1 and 2. Regressions also include indicators of region of residence.

* Asterisks identify estimates that are significantly different from zero at the five-percent level.

References

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.

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

Central Statistical Authority (CSA), Federal Democratic Republic of Ethiopia (December 1996), The 1995/96 Household Income, Consumption and Expenditure Survey (Volume I).

Central Statistical Authority (CSA), Federal Democratic Republic of Ethiopia (November 1999), Report on the 1998 Welfare Monitoring Survey, Statistical Bulletin 224.

Dercon, Stefan (2002), "Poverty and Deprivation in Ethiopia," report prepared for the World Bank.

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.

Schaffner, J. (2002a), "Heteroskedastic Sample Selection and Developing Country Wage Equations," Review of Economics and Statistics 84(2), pp.269–80.

Schaffner, J. (2002b), "Literacy and Schooling in Ethiopia," processed.

Schaffner, J. and S. Banki (2002), "'Cleaned' EMIS92 Data Documentation," processed.

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".