

HOUSEHOLD CONSTRAINTS ON SCHOOLING BY GENDER: EMPIRICAL EVIDENCE FROM ETHIOPIA¹

Pauline Rose and Samer Al-Samarrai²

IDS Working Paper 56

Summary

Using data collected during school-based surveys in two regions in Ethiopia, bivariate and multivariate analysis has been carried out to identify some of the important constraints to primary schooling. Problems in obtaining a representative sample of children not in school for this type of analysis are identified, and allowances are made for these problems, where possible. The results provide some evidence that well-nourished children from wealthier households whose mother and father are literate are more likely to attend and complete primary school. Furthermore, starting school at the official starting age of seven years and not repeating a grade increases the chances of completion. For children living within the vicinity of the school, opportunity costs do not appear to have a significant impact on the probability of attending or completing school, although girls in and out of school have a heavier work burden, relative to boys. Whilst similar factors are found to influence the probability of both boys and girls attending and completing school in the sample, their impact on girls is greater – undernourished girls from households lacking possessions, whose parents are illiterate, are 20 per cent less likely to attend school and 16 per cent less likely to complete school compared to a boy with these characteristics. Starting school above the official age and repeating a grade are also shown to reduce the chances of girls completing school by a slightly greater amount than boys. Statistical analyses of the kind used in this paper are not able to control adequately for cultural constraints to schooling, although these are recognised as potentially providing an additional important explanation for the gender gap.

¹ The authors would like to thank all members of the FAWE Gender and Primary Schooling in Africa programme and Barry Reilly for assistance with this paper.

² Pauline Rose and Samer Al-Samarrai are Research Officers at IDS.

Preface

The work presented in this paper is part of a research Programme on Gender and Primary Schooling in Africa which is being conducted in three African countries under the auspices of the Forum for African Women educationalists (FAWE). The main aims of the project have been to examine the causes of low enrolments, persistence and performance of girls relative to boys and to identify the most promising policy options to facilitate the universal enrolment of children in primary schools.

Christopher Colclough
Programme Director

1 INTRODUCTION

The paper seeks to identify some of the important individual and household characteristics determining the probability of a child attending and completing primary school in Ethiopia, and to see how these characteristics differ by gender. The data set used in this paper was collected during fieldwork carried out in Ethiopia in 1995, conducted as part of the Gender and Primary Schooling in Africa Programme. In order to examine the causes of non enrolment, drop-out and poor performance of girls relative to boys, surveys were carried out in two areas within Ethiopia using a combination of quantitative and qualitative research tools. The survey work was not intended to be representative, in a statistical sense, of the country as a whole. Our research design was intended to complement the results of other studies carried out in the country and to achieve a fuller understanding of the differences in causal factors, across areas with different characteristics.³ This paper concentrates on the use of bivariate and multivariate analysis to examine some of the individual and household characteristics which have a significant impact on the probability of a boy or girl attending/completing primary school in the two survey areas.

The focus on individual and household characteristics is not intended to gainsay the importance of cultural and school-related constraints to schooling – indeed, these play a substantial role in determining whether or not a child will attend/complete school and provide important explanations for the gender gap in enrolment.⁴ These factors (including, for example, early marriage, fear of pregnancy, lack of school places and gender insensitivity of the learning environment) are discussed in detail in Rose *et al.* (1997). This paper concentrates on the economic constraints on schooling which have been identified by the Gender and Primary Schooling study, and by other studies in Ethiopia as being amongst the most important determinants of non-enrolment and drop-out (Rose *et al.* 1997; World Bank 1996a; World Bank 1996b; Yelfign *et al.* 1995; USAID 1994; Asseffa 1991 and Anbesu and Junge 1988). These constraints are amenable to direct policy interventions (for example, direct costs can be reduced by removing school fees or introducing subsidies; opportunity costs can be reduced through more flexible timetabling and/or reducing the school starting age).⁵ It is, therefore, necessary to assess the extent of the impact of economic constraints on

³ A more detailed discussion of the survey sites and an analysis of the factors affecting non-enrolment, drop-out and poor performance from the descriptive statistics and qualitative information can be found in *Gender and Primary Schooling in Ethiopia*, Rose *et al.* (1997).

⁴ The concept of gender is itself based on an historical and cultural understanding which is socially constructed rather than biologically determined.

⁵ It is recognised, however, that policies aimed at tackling direct and opportunity costs alone, without addressing social and cultural barriers to girls' schooling in particular, are unlikely to be effective. Sensitisation of the community and teachers, for example, is extremely important to ensure that parents perceive that girls' education is beneficial, and that teachers provide girls with the appropriate learning environment to encourage them to continue with their schooling.

schooling and, given that girls' access to schooling is considerably lower than boys, to evaluate whether they provide an explanation for the gender gap in enrolment.

The paper begins with a brief background to the economic and educational context in Ethiopia. This is followed by a review of previous studies examining the reasons for non-enrolment and drop-out in the country. Section 4 provides background information on the data and survey areas; and Section 5 describes the type of bivariate and multivariate analysis undertaken, and notes some of the problems of using the data set for such analysis. The results are reported and discussed in Section 6.

2 ECONOMIC AND EDUCATIONAL CONTEXT

The population of Ethiopia is the second largest in Africa and the country is extremely diverse in terms of language and ethnicity. According to the 1994 census, the population was estimated to be 52.3 million, growing at approximately 2.6 per cent per annum (World Bank 1996c). In 1994, approximately 17 per cent of the population was of primary school age (7–12 years). Given the present high fertility rate (7.7 children per woman), the age structure of the population will be slow to change.

Ethiopia is amongst the poorest countries in the world. According to the World Development Report, GDP per capita in 1993/94 was US\$100, the third lowest in the world. The country's social indicators are also low relative to other countries in sub-Saharan Africa. The 1992 estimates for life expectancy at birth (48 years), infant mortality (119 deaths per 1,000 live births), and adult literacy rate (33 per cent) compare unfavourably with those for sub-Saharan Africa (United Nations 1995). According to the 1995 United Nations Human Development Report, Ethiopia ranks among the lowest four countries in the world in terms of social development.

The country is currently in the process of social and economic transition. The long and devastating civil war ended in 1991, and the transitional government (replaced by the Federal Democratic Government of Ethiopia in 1995) took over from the previous socialist regime, known as the *Derg*. The command economy of the *Derg* is now being replaced by a market-led system with increased private sector participation. Since 1992, the country has undertaken economic reforms in conjunction with its first structural adjustment programme. However, the economy remains dependent on agriculture which is mainly rain-fed. As a result, agricultural production is extremely volatile, resulting in erratic GDP growth rates.

Ethiopia is one of the most educationally disadvantaged countries in the world. The primary gross enrolment ratio in 1993/4 of 24 per cent is considerably lower than the average for sub-Saharan Africa (73 per cent in that year). Furthermore, the primary enrolment ratio for girls is considerably lower than that for boys (19 per cent and 30 per cent, respectively). If these rates remained unchanged, the absolute number of children out of school would double by the year 2015 as a result of the projected growth in the school-aged population.

There are favourable indications that school enrolments have been increasing in recent years. This is partly due to the restoration of peace and order, and to the commitment of the new government towards improving education and other social sectors. It is apparent, however, that the increase in enrolments is benefiting boys more than girls. In 1993/94, girls' enrolment was approximately 38 per cent of total enrolment in primary schools compared with 42 per cent in 1991/92. Projections for enrolments over the next five years suggest that, under present conditions, the gender disparity would worsen as total enrolment increases, resulting in girls' enrolment falling to 34 per cent of the total by 1998/99 (Rose *et al.* 1997). The national figures mask regional and rural/urban differences in the gender gap. For example, in 1994/95, girls' proportion of total enrolment varied from 52 per cent in Addis Ababa to 29 per cent in South Ethiopia and Somali; and from 47 per cent in urban areas to 31 per cent in rural areas.

The low proportion of children, especially girls, in school highlights the need to identify the factors influencing non-attendance and non-completion, in order to seek strategies to ensure that primary schooling for all can be achieved.

3 REVIEW OF THE LITERATURE

Three school-based studies have been conducted by the Ministry of Education, in different parts of the country, which sought to identify the factors affecting female participation and performance in rural primary schools in Ethiopia (Yelfign *et al.* 1995; Asseffa 1991; Anbesu and Junge 1988). Yelfign *et al.*'s study was carried out around six schools in Cheha *wereda* (district) in the densely populated Gurage zone. Asseffa's study covered one *awraja* in each of four regions (according to the previous regional boundaries). The areas were selected on the basis of economic diversity: subsistence farming in Wollega; cash-crop farming in Illubabor and surplus-producing farming in Arsi and Shoa. The study by Anbesu and Junge was concentrated in Bahir Dar (which was an *awraja* at that time, but is now the capital city of West Gojjam zone in the Amhara region) and covered 33 primary schools. Data were collected from questionnaires and interviews with pupils, drop-outs, teachers, School Directors and parents for these surveys and the results were analysed using descriptive

statistics. In addition to these surveys, USAID has conducted a larger study of the *Demand for Primary Schooling in Rural Ethiopia* (USAID 1994). The study was carried out in four areas of the country (Bale in the Oromiya region, Welaita in South Ethiopia, Semen Gondar in the Amhara region, and Tigray) and covered ten *weredas* in each. Information was collected from questionnaires with households, primary school pupils and interviews with School Directors, as well as from focus group discussions held with parents, and girls who had never attended school.

Multivariate analysis has been carried out using data from two household surveys to investigate the household demand for schooling in rural Ethiopia (World Bank 1996a); World Bank 1996b). One of these surveys uses data collected from 15,000 households in 15 villages in Regions 1, 3, 4 and South Ethiopia as part of the Ethiopia Rural Household Survey, administered by the Department of Economics at Addis Ababa University in collaboration with the Centre for the Study of African Economies at Oxford University. The data for the other survey were collected from over 7,000 households by the Central Statistical Authority in their 1995/96 Household Income, Consumption and Expenditure, and Welfare Monitoring Survey.⁶

All of the studies identified economic constraints as an important reason for non-attendance and drop-out. Direct costs to schooling are one aspect of this. The World Bank (1996a) and USAID studies both found that the cost of clothing was the most inhibiting school expense. However, neither study found that these costs were significantly different for boys and girls.⁷ Families of higher socio-economic status (estimated by ownership of a radio, metal bed, latrine or barn) were found by the USAID study to be more able to send their children to school. The importance of the costs of schooling is reinforced by the regression results – the World Bank studies show that a proxy for household income has a positive and significant effect on child attendance and completion.

⁶ These papers were prepared for the World Bank's PHRD study. Both papers analysed the probability of currently being enrolled using a logit regression. The World Bank 1996b study also analysed the probability of completing school. Neither of the studies provided separate results by gender, although the coefficients on the gender dummy, in each of the equations, suggest that the probability of girls being currently enrolled is significantly lower than of boys.

⁷ There is a large difference in the estimate of the direct costs of schooling provided by these two studies. Including clothing and stationery supplies, the USAID estimate is double the estimate in the World Bank paper (or 1.7 times higher including only school stationery). Both imply, however, that school expenses are likely to place a heavy burden on parents, especially in families where there are several children of school-going age.

In group discussions with parents in the USAID study, hunger or insufficient food due to drought were frequently mentioned as reasons for not enrolling children in school. Data for nutritional status are not, however, available in any of the studies.

Parental education is consistently found to influence whether or not a child attends school. This is both due to the relationship between education and socio-economic status, as well as the influence of parental education on their attitudes towards schooling. Attitudes are particularly important because parents in Ethiopia perceive that the returns to schooling are low (especially for girls), partly because the benefits are seen in relation to the opportunities for waged employment which are in short supply (USAID 1994; World Bank 1996a). The results of Yelfign *et al.*'s study suggest that girls in particular are more likely to attend school if their mother is literate. The regression results show that educated fathers and mothers (or household head) have a strong positive effect on attendance and completion.

Indirect costs are also considered to be a constraint on school attendance and completion. The USAID study found that there was a conflict between schooling and seasonal demands on the household. Gender differences in enrolment were mainly attributed to higher opportunity costs of girls' time: in the USAID study, girls out of school reported spending an average of 14–16 hours a day on a variety of tasks, leaving them no time to attend school. In the papers using econometrics, household size was included as a proxy for the indirect costs of schooling: it was assumed that larger households may not need the contribution of the labour of all children in the household.⁸ The positive impact of larger households upon the probability of attending school, in these regressions, was interpreted in this way.

If children live with both parents there may be more support for their schooling, and this characteristic is shown by the World Bank (1996b) paper to be significantly associated with higher rates of enrolment and completion. The World Bank (1996a) study hypothesises that children who live in female-headed households are more likely to attend because such households are less likely to own land and, thus, the opportunity cost of schooling is reduced.⁹ The results are, however, not significant in that regression. According to the results of the World Bank (1996b) study, children in female-headed households are, however, significantly more likely to be enrolled in, and complete, school than those in male-headed households. The

⁸ It is possible that larger household size negatively influences the probability of a girl attending school because of the increase in the amount of work within the household, for example looking after siblings and preparing food. However, the results of the regressions are not analysed by gender.

⁹ There may, however, be an opposite tendency; that mothers in female-headed households may have to earn an income, so daughters are forced to stay at home and substitute for their mothers in the household.

paper proposes that this is because mothers have a more positive attitude towards schooling. In contrast, those living with guardians other than parents or grandparents are found by the World Bank (1996a) study to be less likely to go to school, which the paper suggests is because of discrimination against children who are not direct descendants, or because other members of those households are less likely to be children of school age (increasing the opportunity costs of schooling).

Anbesu and Junge's study indicates that in Bahir Dar there were more overage boys than girls in school and more underage girls than boys in school. They suggest that overage children are more likely to drop out. Girls starting at an early age were particularly likely to have a better chance of completing, which Anbesu and Junge propose is probably related to puberty.¹⁰ The results of two regressions on the probability of being currently enrolled in school provide different results – the World Bank (1996b) paper shows that age is negatively related to enrolment; whereas World Bank (1996a) indicates that children are more likely to be enrolled the older they are up until a certain age, after which the probability of enrolment falls. Both suggest, however, that older children are less likely to be enrolled because of the higher demand for their labour.

World Bank (1996a) finds that children from Muslim families are significantly less likely to be enrolled than those from other religious backgrounds, suggesting that their parents have a negative attitude towards schooling.¹¹ The regressions do not, however, provide any evidence of cultural influences on the probability of attending and completing school, although these have been shown by other studies to be important. Early marriage was identified by a number of such studies as being a constraint on schooling, especially for girls (Anbesu and Junge 1988; Asseffa 1991; USAID 1994). The extent of the problem differs across the country with one study proposing that early marriage was not a problem in the district surveyed (Yelfign *et al.* 1995), whilst another suggests that parents believed that early marriage was a less important constraint than economic factors, although it was seen as an important consideration in deciding between whether to send a boy or girl to school (USAID 1994).

School quality has been shown in a number of the studies to have only limited influence on enrolment. For example, the USAID study found that there was a weak relationship between school expenditure on stationery and learning materials and school attendance. This was

¹⁰ The later girls start, the more likely they will have reached puberty before completing, which puts pressure on them to leave school and marry.

¹¹ Since some areas are predominantly Muslim this variable could also be acting as a proxy for other factors, such as ethnicity.

suggested to be due to very few parents having been to school themselves and were, therefore, unaware of issues related to school quality. Teacher characteristics appeared to have an influence on demand: teachers with longer teaching experience had a positive effect on demand. Overall, the study suggested a positive relationship between girls' enrolment and the proportion of female teachers in the schools. Contrary to expectations, the World Bank (1996a) study found that the number of pupils per teacher and shortage of books were both positively related to being enrolled, suggesting that there was high demand for schooling in particular areas, and that supply was not responding to this.¹²

Another reason proposed by Anbesu and Junge, and by the USAID study, for non-attendance was distance to school. In both the surveys, the majority of those in school lived within 30 minutes walk of the school, suggesting that school attendance was biased by ease of accessibility.¹³ In both of the papers using multivariate analysis, however, distance to school was not statistically significant.

Thus, the results of previous studies suggest that children are more likely to be enrolled in primary school the wealthier the household; the better the nutritional status of the child; the more educated are their parents; the larger the household size; and if they live with both parents. Children are more likely to complete primary school for similar reasons, as well as if they start school at an early age. Although distance to school is considered to be an important factor, the data do not provide clear evidence on this; and school-related factors are not found to have a strong effect on the demand for schooling. The main reasons provided for the lower enrolment of girls relative to boys were related to higher opportunity costs of girls' time and early marriage. The influence on the constraints to schooling for girls relative to boys in two different areas of the country are investigated in more detail below.

4 DESCRIPTION OF THE DATA

Data for this paper were collected during fieldwork carried out in East Gojjam zone (Amhara region) and Jimma zone (Oromiya region). These two zones were picked due to their differing characteristics in terms of gross enrolment ratios, females as a proportion of total primary enrolment and variations in socio-economic conditions (e.g. main economic activity, ethnicity and religion). The gross enrolment ratio is below the national average in both regions (Table 1), and is higher in the Oromiya region than the Amhara region. The gender gap is wider than

¹² School-specific characteristics are not included in the World Bank (1996b) regressions.

¹³ Distance is also likely to increase the opportunity costs of schooling as more time is spent travelling to and from school.

the national average in the Oromiya region, whereas it is narrower than the national average in the Amhara region.¹⁴

Table 1: Gross enrolment ratio in grades 1–6, 1994/95

	Male GER	Female GER	Total GER
Amhara	20.7	18.0	19.3
Oromiya	30.2	15.6	23.1
National average	35.7	22.1	29.0

Source: Ministry of Education (1996).

In East Gojjam the population is fairly homogenous in terms of ethnicity, language and religion: the majority of the population is Amharic-speaking Amharas and the main religion is Orthodox Christianity. In contrast, Jimma zone is relatively heterogeneous, although the majority of the population is Oromigna-speaking Oromos, and the main religion is Islam. In both zones over 80 per cent of the population live in rural areas and the major economic activity is farming. Jimma zone is predominantly a cash crop area (growing coffee and *chat*), although there is also some cereal crop farming. Subsistence cereal crop farming (including *teff*, maize, wheat, oil seed and barley) is the main activity in Est Gojjam, and farmers also keep livestock.

Within each of the zones, two *weredas* (smaller administrative units) were chosen to represent the characteristics of the zones: Manaa and Saqqa Coqorsaa in Jimma zone and Machakel and Gozammin *weredas* in East Gojjam zone.¹⁵ Within each of the zones both town and rural schools were visited in order to obtain a picture of the different demand and supply characteristics of schooling in these areas. Since the majority of the population in both zones live in rural areas a total of two urban and nine rural schools were surveyed. Table 2 provides some of the school characteristics of the zones and *weredas* surveyed.

Table 2: Schools and enrolments in survey areas 1994/95

	Number of schools			Primary enrolment	
				Rural	Urban

¹⁴ There is a similar pattern in the gender gap between the zones (36 per cent of those enrolled in primary school are female in Jimma; compared with 43 per cent in East Gojjam), but enrolment ratios are not available by zone.

¹⁵ An additional criterion was imposed by the fact that there were very few rural schools in either of the zones visited with grade six pupils, so *weredas* had to be selected according to the availability of grade six pupils.

<i>Wereda</i>	Primary	Junior Secondary	Senior Secondary	Total	% female	% grade 1	Total	% female	% grade 1
Gozammin*	16	1	0	2561	42	68	-	-	-
Machakel	19	2	1	2427	34	74	1047	52	29
Sinan**	20	1	0	2457	25	81	589	43	24
East Gojjam	346	34	8	36275	40	76	22871	49	31
Maanaa	18	1	1	3271	39	63	878	48	48
Saqqa Coqorsaa	35	3	1	4965	33	63	1560	46	37
Jimma	303	24	8	37739	30	64	26615	46	31

* Gozammin is the *wereda* surrounding Debre Markos town and does not have an urban centre (Debre Markos is a separate *wereda* and is entirely urban).

** Sinan was a separate *wereda* in 1994/95 but was divided between Machakel and Gozammin in 1995/96.

Source: Rose *et al.* (1997).

Due to the recent increases in enrolments and the high drop-out rates after the first year of primary school, a large proportion of enrolment is in grade one.¹⁶ On average, girls as a proportion of total enrolment is higher in East Gojjam than in Jimma, and higher in urban areas compared to rural areas.

The data used in this paper were collected from three sets of interviews with grade six pupils (the last grade or primary), drop-outs and non-enrolled children.¹⁷ In the rural school areas, all grade six pupils, together with as many drop-outs and non-enrolees in the same age group as grade six pupils that could be identified, were interviewed.¹⁸ The same procedure was used in the urban areas, except that a selective random sample of grade six pupils was taken, reflecting the distribution of male and female pupils in this grade.¹⁹

¹⁶ The sample of drop-outs reflects this pattern: of those interviewed, 60 per cent in Jimma and 46 per cent in East Gojjam only reached grade one or two.

¹⁷ The fieldwork also included interviews with educational officials, School Directors and teachers; and group discussions with School Management Committees, parents and pupils. These are discussed in Rose *et al.* (1997).

¹⁸ Officially, grade six pupils should be 12 years old. In practice, however, ages range from below 12 to over 20 due to late entry, repetition and re-entry. The age range of drop-outs and non-enrolees also reflects this dispersion.

¹⁹ "Urban" is defined by the government as a locality with 2,000 or more inhabitants. All administrative capitals (regional, *awraja* and *wereda*) and localities in which urban dwellers' associations were established are also considered as urban centres, regardless of their population size (DSA 1991). This definition is also used in the collection of education statistics. The urban areas referred to in this paper

Table 3: List of variables

Variable name	Description
School dummies	Reference category: if the child is from Emmanuel school area
Buxeri Gabissa	1 if the child is from Buxeri Gabissa school area
Somodo	1 if the child is from Somodo school area
Olmee Yaboo	1 if the child is from Olmee Yaboo school area
Coqorsaa	1 if the child is from Coqorsaa school area
Yebbuu	1 if the child is from Yebbuu school area
Embole	1 if the child is from Embole school area
Karar	1 if the child is from Karar school area
Yeted	1 if the child is from Yeted school area
Inarate	1 if the child is from Inarate school area
Kabe	1 if the child is from Kabe school area
Education dummies	Reference category: if father/mother is illiterate
Father formal	1 if the child's father has some formal education
Father non-formal	1 if the child's father has some non-formal education
Mother formal	1 if the child's mother has some formal education
Mother non-formal	1 if the child's mother has some non-formal education
Guardian dummies	Reference category: if children live with both parents
Mother only	1 if the child lives with their mother only
Other guardians	1 if the child does not live with their mother only or both parents
Other variables	
Distance	Distance in km of child's home from school
BM15	1 if the child is below the fifth percentile of the body mass index for age distribution (i.e. the child is under-nourished)
Work for household	Composite index of the amount of work the child does for the household
Wealth index	Composite index of possessions of the child's household
No of children	Number of children in child's household
No of adults	Number of adults in child's household
Repeated	1 if the child has repeated one or more grades of primary school
Starting age	Age at which the child started primary school

Interviews with pupils, drop-outs and non-enrolees in each of the survey area collected information on a number of child and household characteristics (including sex, age, school starting age, weight, height, distance to the nearest school, whether the child has repeated a grade, work performed by the child for the household, household possessions, level of education and occupation of mother and father, guardians, and number of children and adults in the household). A full list of the variables used in the present analysis is shown in Table 3.

5 ANALYSING THE DATA

are small towns with similar characteristics to the surrounding rural areas, except that they are more densely populated.

5.1 Bivariate analysis

Bivariate analysis was carried out on the data collected from the questionnaires. The analysis examines differences in characteristics between boys and girls within each of the surveys; and differences in characteristics between the three surveys (i.e. pupils, drop-outs and non-enrolees) for girls and boys separately. Tests have been carried out to see whether the differences in the proportions or means between groups of data are significant (i.e. between boys and girls within surveys; and between surveys). Two tests are used, one for the categorical variables in the data set and another for the continuous variables. For the categorical variables, the proportion in one group is tested against the proportion in the other group (for example, the proportion of male pupils living with both parents is compared with the proportion of female pupils living with both parents) to see whether the proportions are significantly different. For a continuous variable (such as school starting age), a test of the difference in means is used to see whether there is a significant difference between the two groups.²⁰

For the difference between the sexes within each of the surveys, a first round of analysis was undertaken looking at both regions separately. On the whole, the same relationships were found to hold between boys and girls in each region. Tests of significance between the sexes are, therefore, reported for the two regions combined (see Appendix 1). Analysing the differences between the surveys, it was found that there are significant variations between the regions in some of the child and household characteristics. Thus, the tests of significance between the surveys are reported by region (see Appendix 2).

5.2 Multivariate analysis

In addition to the bivariate analysis, two logistic regressions were run.²¹ The purpose of running these regressions was to examine the differential impact of household attributes on boys' and girls' education. In each of the two regressions a set of school dummies were included to control for school and local characteristics (for example, ethnicity and religion).²² The dependent variable in the first regression was whether a child had been to school or not.

²⁰ For a summary of the tests used see Newbold (1984: 363 and 368). The test between the proportions tests the equality of two population proportions and is valid for samples with approximately 40 observations or more. The test between the means tests for differences between population means and assumes that the samples are drawn independently. The test is valid for samples with approximately 30 observations or more.

²¹ For a complete description of the uses of limited dependent variables and qualitative variables in econometrics see Maddala (1983). For a summary of the methodology used in this paper see Al Samarrai and Peasgood (1997).

²² The sample broadly reflects the characteristics of the school areas in terms of ethnicity and religion (see Tables 3.11, 3.12 and 3.13 in Rose *et al.* 1997).

In this case, both drop-outs and pupils were compared with non-enrolees. The dependent variable in the second regression was whether a child had reached grade six or not, conditional on having attended school.²³ In this regression the drop-out sample was compared against the pupil sample.

For the regression analysis, the boys' and girls' samples were pooled.²⁴ In order to assess the differential impact of household attributes on boys' and girls' education, a set of gender slope dummies (equal to one if the child is female, zero otherwise) were included on some of the coefficients. Initially gender slope dummies were included for most of the explanatory variables included in the regression analysis. Some of the gender slope dummies were subsequently dropped due to their insignificance and a more parsimonious model was estimated for each regression.

5.3 Problems with the data set

Table 4 shows the number of children interviewed in each group by school and zone and highlights some of the problems of using the data set for comparative analysis between pupils, drop-outs and non-enrolees. The proportions of the data in each group shown in Table 4 are not equal across the schools or across the zones; and the relative sizes of the three samples in each school area are not representative of the proportions in these groups. For example, the gross enrolment ratio in primary school for Amhara is 19.3 per cent (see Table 1). This implies that only 19.3 per cent of children of primary school-going age are in primary school whereas 40 per cent of the sample for East Gojjam (in the Amhara region) are attending primary school.²⁵ This suggests that pupils in the survey are over sampled compared with drop-outs and non-enrolees.²⁶ This is not a problem for the bivariate analysis as it implies that the pupil proportions and means estimated from the sample will be more accurate compared to the drop-out and non-enrollee estimates. It might, however, be a problem in the multivariate analysis, since the relative proportions of those going to school and completing school are not representative of the population as a whole.

Table 4: Number of children sampled in each group by school and zone

²³ In the rest of the paper those children that have reached grade six are referred to as completers.

²⁴ A likelihood ratio test was carried out to test whether the pooled regression was a better specification than using two gender disaggregated regressions. The test was not rejected implying that the pooled regression provided a better specification.

²⁵ This comparison is only meant to be indicative. The pupils in the survey are from the last grade of primary so the enrolment ratio within this group is likely to be considerably lower than 19.3 per cent given the high rates of drop-out between grades one and two.

²⁶ Data on drop-out rates or the proportion of non-enrolees are not available for the regions as a whole or for the zones in which the survey took place. Therefore, it is unclear whether the proportions of drop-outs compared to non-enrolees is representative in the survey.

School/zone	Number of pupils surveyed	Number of drop-outs surveyed	Number of non-enrolees surveyed
Jimma			
Buxeri Gabissa (r)	8	1	11
Somodo (r)	13	14	10
Olmee Yaboo (r)	23	11	22
Coqorsaa (r)	21	2	24
Yebbuu (u)	51	19	19
Jimma total	116	47	86
East Gojjam			
Embole (r)	10	11	9
Karar (r)	7	11	14
Yeted (r)	11	4	7
Inarate (r)	10	12	14
Kabe (r)	8	3	10
Emmanuel (u)	51	27	20
East Gojjam total	97	68	74

r = rural
u = urban

The number of children out of school in each school area is not indicative of the total number of children out of school in that area.²⁷ For example, although many more children out of school were interviewed in Yebbuu than in Buxeri Gabissa this does not imply that the proportion of children out of school in Yebbuu is much higher than in Buxeri Gabissa. One implication of this is that certain school areas have more weight in the analysis than others and hence estimates from the data will be biased towards these schools. For both zones, the urban schools tend to have large sample sizes in each of the three groups of children, implying that the parameter estimates will be biased towards the urban schools. In the multivariate analysis more weight will thus be given to the urban schools where the relationship between household characteristics and the dependent variable may be different to that in rural areas. This is tested using the likelihood ratio test which involves testing the pooled regression against separate regressions for the rural and urban samples.²⁸ The restriction implied by pooling the data in the first regression (where the dependent variable is whether the child attends school) is not rejected. This implies that the relationship between school attendance and the explanatory variables used are very similar in urban and rural areas.

²⁷ Areas where more children out of school were interviewed are school areas where it was possible to identify these children. They were easier to identify in the urban school areas, which is reflected in the data reported in Table 3.

²⁸ For details of the test used see Greene (1993: 647).

Testing the second regression (for the probability of completing primary school), the restriction implied by pooling the rural and urban samples is rejected in favour of running separate regressions for the rural and urban samples.

For the multivariate analysis, the samples from the two zones are pooled. As mentioned previously, the two zones were chosen to represent very different characteristics and, therefore, the relationship between household characteristics and the dependent variable may be different in the different zones. A likelihood ratio test to test the pooling assumption is again not rejected in the first regression but is in the second regression.²⁹

Schools often serve a wide area, and many non-enrolees and drop-outs live in remote areas making it difficult to interview these children. Most of the drop-outs and non-enrolees interviewed were, therefore, living in the vicinity of the school. The mean distance children lived from the nearest school was smaller for the children out of school than for the grade six pupils. This implies that the children out of school in the sample are a specific sub-set of the population of children out of school and this must be borne in mind when looking at the results. An attempt was made to account for this in the analysis by restricting the sample to those children living within a four kilometre radius of the school. This reduced the pupil sample by more than the drop-out and non-enrollee samples. Therefore, the sample of children used for the analysis can be interpreted as a sample of children living within reasonable walking distance from the nearest school.³⁰

The problems outlined in this section highlight the difficulties in trying to collect information on a representative sample of children in one area where a recent and accurate sampling frame is not available. Schools, particularly in rural areas, serve a wide area and reaching children who do not go to school is difficult because households in which these children live are widely dispersed. Furthermore, neither schools nor Peasant Association offices have records of children who are not in school (either drop-outs or non-enrolees), and maps identifying the location of households in the area do not exist.

²⁹ A second test was carried out to test the restriction of pooling against the alternative of running separate regressions for each *wereda*. For both of the regressions the pooling restriction was rejected in favour of running four separate regressions for each *wereda*.

³⁰ Given that large distances to school have been cited as a major obstacle to school attendance (see Section 2) it is likely that the gross enrolment rate for children living within four kilometres of the school is much higher than for the region as a whole. Therefore the proportions of pupils and out of school children in our sample may be more representative when this restriction is applied. Unfortunately, no comparative regional data are available.

Due to the problems mentioned in this section the multivariate analysis must be interpreted with caution. The likelihood ratio tests suggest that the completion regression may be more problematic than the attendance regression. However, the bivariate analysis is generally supportive of the econometric analysis and, hence, both are reported below.