DETERMINANTS OF SCHOOL ENROLLMENT IN ETHIOPIA

BY
MULUGETA GEBRESELASSIE

A THESIS PRESENTED TO
THE SCHOOL OF GRADUATE STUDIES
ADDIS ABABA UNIVERSITY
IN (PART) FULFILLMENT FOR
THE DEGREE OF MASTER OF SCIENCE IN
STATISTICS

ADDIS ABABA
JUNE 1998

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

DETERMINANTS OF SCHOOL ENROLLMENT IN ETHIOPIA

BY
MULUGETA GEBRESELASSIE
DEPARTMENT OF STATISTICS
FACULTY OF SCIENCE

Approved by the Examining Board:

Dr. Deepak Grover, Examiner ______________

Dr. Tadiwos Koroto, Examiner ______________

Dr. Abebe Tessera, Advisor ______________

ACKNOWLDGEMENT

The highest credit goes to my advisor Dr. Abebe Tessera , who has been very grateful from the beginning to the end of the thesis work. All his comments were valuable. He taught me curiosity of work for my life.

I can not forget the help and the day-to-day follow up I got during my thesis writing from Amare Asgedom. He was very helpful to read and comment my literature in addition to his personal encouragement. I am also indebted to my friend Abeba Abraha, who has been very helpful during my job. The encouragement I enjoyed from my friends Assefa F., Mulugeta M. and Asmelash A. is worth mentioning.

Finally my thanks goes to my family, my mother, my sisters and my brothers who were tolerant and helpful since my early college days and to the DAAD who sponsored my graduate education.

ADDIS ABABA UNIVERSITY LIBRARIES

ABSTRACT

This paper assesses empirically the role of household, community, individual and regional factors in determining school enrollment in Ethiopia. It uses logistic regression and discriminant analyses methods to analyze the data. The analyses is done in such a way that first determinants of primary and secondary school enrollment are studied and then the overall school enrollment is studied. Finally appropriate diagnosis is done for each of the models.

The results from the analyses showed that age and sex of the person, relation of the person to the head and spouse (common child or not?), location of residence (urban/rural), education and occupation of the head and the spouse, the interaction of sex and location of residence and the interaction of sex and distance to primary school , were important determinants of school enrollment. Distance to school, number of school age people in a household, sex of the head, proportion of girls in the household, the logarithm of school expenditure, return's from education and the interaction of education and occupation of the head were found important determinants of primary schooling in addition to those listed above. Moreover, secondary schooling is found to be a function of education of the head, urban rural residence and size of school age people.

In general, the results of data analysis demonstrate the usefulness of applying multivariate techniques to the analysis of inter correlated variables. The techniques prove to be effective to the extent at least 72% in discrminant analysis and up to 83% in logistic regression of the students (cases) are correctly classified by the computed functions.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENT ii

ABSTRACT iii

CHAPTER ONE **INTRODUCTION** 1

CHAPTER TWO **LITERATURE REVIEW** 4

2.1 GENERAL BACKGROUND 4

2.2 OTHER COUNTRIES EXPERIENCES 5

2.3 STUDIES ON ETHIOPIA 9

2.4 SHORTCOMINGS OF THE REVIEWED STUDIES 14

CHAPTER THREE **OBJECTIVES OF THE STUDY, DATA AND METHODOLOGY USED** 16

3.1 OBJECTIVES 16

3.2 THE DATA 17

3.3 VARIABLES IN THE STUDY 20

3.4 METHODOLOGY 24

3.4.1 LOGISTIC REGRESSION 25

3.4.1.1 STATISTICAL INFERENCE 26

3.4.1.2 MODEL CHECKING 27

3.4.2 DISCRIMINANT ANALYSIS 29

CHAPTER FOUR. **RESULTS OF DATA ANALYSES**............................32

4.1 LOGISTIC REGRESSION RESULTS..................................33

4.2 DISCRIMINANT ANALYSIS RESULTS............................38

CHAPTER FIVE. **DISCUSSION, CONCLUSION AND RECOMMENDATION**..48

REFERENCES..53

ANNEXES..56

Table 1. DETERMINANTS OF MALE ENROLLMENT

Table 2. DETERMINANTS OF FEMALE ENROLLMENT

Table 3. DETERMINANTS OF URBAN ENROLLMENT

Table 4. DETERMINANTS OF RURAL ENROLLMENT

Table 5. DETERMINANTS OF ENROLLMENT FOR URBAN MALE

Table 6. DETERMINANTS OF ENROLLMENT FOR RURAL MALE

Table 7. DETERMINANTS OF ENROLLMENT FOR URBAN FEMALE

Table 8. DETERMINANTS OF ENROLLMENT FOR RURAL FEMALE

Table 9. SUMMARY TABLE FOR URBAN

Table 10. SUMMARY TABLE FOR RURAL

Table 11. SUMMARY TABLE FOR URBAN MALE

Table 12. SUMMARY TABLE FOR RURAL MALE

Table 13. SUMMARY TABLE FOR URBAN FEMALE

Table 14. SUMMARY TABLE FOR RURAL FEMALE

Table 15. MALE -FEMALE COMPARISON

Table 16. WILKS' LAMBDA AND UNIVARIATE F-RATIO FOR ENROL

Table 17. WILKS' LAMBDA AND UNIVARIATE F-RATIO FOR P_ENROL

Table 18. WILKS' LAMBDA AND UNIVARIATE F-RATIO FOR S_ENROL

TABLE 19. FOR ENROL WITH OUT REGIONAL DUMMIES

Table 20. MULTIPLE REGRESSION

TABLE 21. TEST OF EQUALITY OF GROUP COVARIANCE MATRICES USING BOX'S M

FIGURE 1 NPP FOR ENROL

FIGURE 2 NPP FOR P-ENROL

FIGURE 3 NPP FOR S-ENROL

FIGURE 4 SCATTER PLOT OF PREDICTED VALUES AGAINST RESIDUALS

CHAPTER ONE

1. INTRODUCTION

Evidence from published statistics indicate that a strikingly low percentage of school age children in Ethiopia participate in formal education. This participation rate is unacceptably very low even when compared with other low income African countries. For instance, in 1983, the primary gross enrollment ratio (PGER) for Ethiopia was 38%, while the corresponding figure for Sudan was 46.2% (Maglad, 1994) and for the whole Sub-Saharan Africa 75% (The World bank, 1988). This low participation rate has been worsening and in 1994/95 it was 29% (MOE, 1997).

Several reasons have been stated for the low school participation (enrollment). Destafano, et al (1993) give the deterioration of school quality as the main reason for the low demand for schooling. In another study by the same authors (Destafano, et al, 1992), which used parents, teachers and school administrators as its reference, the absence of further educational opportunities that could lead to non-farm employment, the high direct and indirect costs of schooling, irrelevance of schooling to rural life and coerced participation in schooling were mentioned as the main factors that can contribute to this low enrollment rate.

In some developing countries studies have been done to identify socio-economic and demographic determinants of school enrollment. Maglad (1994), for instance in his study of

Sudan found age of child, sex of a child, parents' educational level, land ownership, distance to school (both distance to primary and secondary school), price of schooling (both direct and indirect costs) and region as the most important determinants of school enrollment. In another study done in Brazil, Singh (1992) found household size to be an additional important factor affecting enrollment. However, empirical studies related to determinants of school enrollment in Ethiopia are quite limited. The question, therefore, of the factors that affect school enrollment is of importance.

Accordingly, the aim of this study is two fold. First, to investigate how household, community, individual and regional factors influence child school participation. Second, because of the gender and urban/rural gap in school participation, it will examine the effect of these factors separately for boys and girls, urban and rural and their interaction.

The attainment of universal school participation, especially at primary school level, is prominent on the agenda of educational policy makers. The foremost significance of this thesis is therefore to address policy makers as well as educational researchers with the findings of this study, that is, key determinants of school enrollment in Ethiopia. It will also help to minimize cost in further studies, especially cost of data collection, as the study has summarized the most important factors which determine schooling in Ethiopia.

The study has some limitations. Since the data is obtained from a secondary source, it was very difficult to exhaustively use all the variables that can be found from the data because of missing cases. Moreover, the sample design was so complex that, it could have it own negative influence on the estimates. Especially, the interaction of the complexity of the survey, though it has mild influence on the estimates from large samples, with the missing of some cases could be seen as a threat to the estimates.

The rest of the study proceeds as follows. A review of literature is given in Chapter Two. Chapter Three discusses the general and specific objectives of the study. Description of the study variables and discussion of the methodology are also part of it. Chapter Four deals with the analyses of the data and the discussion of the results. Results of model diagnostics are also given in this chapter. Finally conclusions and recommendations drawn from the study are given in Chapter Five. The Appendix part contains tables which could not be listed in the body of the thesis.

Abbreviations and Definition

CSA	Central Statistics Authority
MOE	Ministry of Education
GER	Gross Enrollment Rate
NER	Net Enrollment Rate
PHRD	Policy and Human Resources Development
AAU	Addis Ababa University
EA	Enumeration Area
Urban	A locality with number of people greater than or equal to 2000. It includes all regional, zonal and wereda capitals.

CHAPTER TWO

2. LITERATURE REVIEW

2.1 GENERAL BACKGROUND

No one would doubt that education is the major vehicle to development and the remedy for people's problems. It is widely accepted that all children should receive at least primary education if human labor is to yield sustainability in the development of a country. But there are many hindrances to popular participation in primary education. In the case of Ethiopia, contrary to the 1990 UNESCO Conference's slogan, "Education for All", access to schooling has been worsening. At less than 20% gross enrollment rate (GER), Ethiopia has the lowest access to schooling in Sub Saharan Africa. However, recent trends are encouraging. For instance Gross Primary Enrollment in 1992/93 was 19.7% and in 1993/94: 23% and in 1994/95: 29% and in 1995/96 it is greater than 35%. To catch up the attainment in the rest of the world, however, there is a lot which remains to be done. To identify the main problems behind this low participation rate, many studies have been made in different countries of the world. However, in the case of Ethiopia the number of studies is quite limited.

2.2 STUDIES ON OTHER COUNTRIES

Tan, Lee and Mingat (1984) studied the determinants of the proportion of eligible siblings enrolled for Malawi. They applied the Ordinary Least Squares regression (OLS) with the dependent variable defined as the proportion of children of eligible ages who are enrolled (number of siblings in school divided by the number of siblings in age interval 5-23 years). They found that cost of schooling (school fees + other schooling expenses) exerts a statistically significant negative affect on the actual number of eligible children that households enroll. Other variables which were statistically significant include, family background characteristics such as mother's education, urban-rural residence and proportion of girls among household's children.

The proportion of children in school was higher in households with better educated mothers, and the effect is reinforced in urban households. Father's income and residence in the northern region were found to be positively associated with high enrollment, where the latter possibly reflects the educational tradition in this part of the country . As expected the proportion of girls was found to be negatively associated with enrollment.

The demand for primary schooling in rural Mali (Birdsall, 1987) was analyzed using OLS regression . The dependent variable was the ratio of the number of persons in the household currently enrolled in school to the number of persons in the household between the ages of 6 and 14. Four sets of independent variables were used: household variables, school quality variables, school fee variables and distance to school variables.

School fees, distance to school and school quality measured by student teacher ratio, number of books per classroom and the late payment of teachers (perhaps reflecting the general level of administrative support teachers can expect) had negative association with enrollment. Among the household variables, the effect of income appeared to be negative, when the ratio of adults to children was included as one of the explanatory variables. This may reflect the contribution of children not in school to income. Ethnic and religious difference across households had no effect on demand once other factors were taken into account. The study also found that school fee is highly negatively correlated with demand than that of distance. If fees were raised and there was no change in the distance or quality of schools (i.e. the funds were used to reduce the overall fiscal burden of the government) the low enrollment rates would fall further. This implies that school fees cannot compensate for quality of schooling by enabling the government construct more schools.

A study in rural Peru (Gertler & Glewwe, 1989), used the logit model for two sets of data, households with local school alternative and households with far away school alternative. And the study showed that level of parents education, presence of other children of secondary school age, sex, school quality and cost of schooling to be the most important determinants of schooling. The result showed that a given increase in user fees will cause about the same proportionate reduction in school enrollment among the poorest (75% of the population) but a much smaller proportionate reduction for the wealthiest (25% of the population). Thus increases in user fees will reduce enrollments less among the richest group when compared to the rest of the rural population in Peru. Besides this, the study found parents' level of education to be positively correlated with enrollment.

The determinants of school enrollment in Sudan (Maglad, 1994) were studied by using a probit model. The dependent variable was the status (enrolled or not-enrolled) of a child and the explanatory variables were age, sex of a child, parent's education level (mother's & father's), land ownership, distance to school (distance to primary and secondary schools), urban -rural residence and region of residence.

The study found families in rural areas to be reluctant to send their daughters to school beyond the primary school level, especially if they have to travel outside their communities, because this raises the worry of parents sending their children (after puberty) to secondary schools in another town . The study confirmed this idea by the fact that constructing 25% more new primary schools would raise boys primary enrollment probability by 10% while that of girls by 8%. But constructing some 25% more new secondary schools in villages more distant from the ones then found increased boys secondary school enrollment by 2.6% and girls secondary enrollment by 5.6%.

The author (Maglad) used distance to school as a measure of cost of schooling. He argues that, the further away the school is from the village, the larger are the costs, since more time is spent in traveling to and from school and might in some cases result in longer periods of absence if the child has to reside at a school boarding or with a relative living near the school. The availability of a school in the village makes it possible to attend school at lower indirect costs, since the child can attend school at times when the demand for his/her labor is at its peak (for example, at harvesting time). Also he found some of the direct costs of schooling, such as transportation costs and costs of meals at school to vary directly with distance to school . In some circumstances transport and meal could represent a major portion of direct costs, especially where no fees are imposed and books are freely provided.

Land ownership was also found to have a significant effect on enrollment. Because land ownership which captures the flow of the non-labor income, that is most individuals especially males, work in their family farms with out payement, could reflect both a price and income effect. This price is the indirect cost of children in rural areas, where child labor contributes considerably to the family's agriculture and livestock production.

The variation on the effect of the different determinants of school enrollments for the different regions in Sudan was studied by using regional dummies in the probit model. The study described regional dummy variables as important to control for some of the unobserved heterogeneity in the community such as quality, ethnic composition and infrastructure in the explanation of child schooling.

As cited in this study on Sudan, Singh (1992) in his study on Brazil , found household size to be an important factor negatively associated with school enrollment. However, Chernichovsky (1985) in a study of schooling in Botswana, found family size to be positively correlated with school enrollment .

2.3 STUDIES ON ETHIOPIA

People can acquire education either through the formal or the non formal way. In Ethiopia, the tradition of formal education dates back to the 6th century B.C. when the Sabean Alphabet was used for instructional purpose (Tekeste Negash, 1990) . But modern formal education started in 1908 with the opening of the Menelik II School in Addis Ababa . In 1930 the number of schools grew to 22, of which 14 were out of Addis Ababa. Starting from this low basement , student enrollment grew 55 fold between 1943 and 1974. The participation rate (enrollment rate), however, remained low. In 1974 the participation rate figures for primary, Junior Secondary & Senior Secondary were only 18, 7 and 3 percent respectively . Distribution of enrollment by gender was not impressive either. In terms of enrollment rate urban areas were better than rural ones (Fasil G. Kiros, 1990).

Destafano, et al (1992), who used parents, teachers and school administrators as their references found the absence of further educational opportunities that could lead to non-farm employment, the high direct and indirect costs of schooling ,irrelevance of schooling to rural life and coerced participation in schooling as the main factors that can contribute to the low demand for schooling in Ethiopia. In another study by the same authors Destafano, et al (1993), deterioration of school quality was found as the main reason for the low demand for schooling. In this study, it is reported that female access to schooling as well as aggregate national figures vary among regions.

A report for the USAID (1993) which used a total of 520 households with school age children from four regional localities (namely Bale, Welaita, South Gondar and Central Tigray) provides a comprehensive study on demand for schooling in rural Ethiopia. The study found that economic constraints represented the most salient impediments to participation and persistence in primary school in the rural areas . In the study both mothers and fathers agreed that opportunity costs (need to work at home) and school costs were the biggest bottlenecks to sending their sons and daughters to school. Those who send their sons to school reported that, they perceive boys as returning parents' investment in schooling. However, in South Gondar, girls were favored for enrollment over boys because parents need boys' labor at home and because boys are apt to wander and require a higher investment, according to some parents. The study concluded that "while school quality is undoubtedly an important factor in the learning process, initially improving school quality in ways that do not alleviate households of financial burdens will probably not prove the answer to increasing enrollments in village schools. Parents must be educated to be both consumers of and informed consumers of schooling in order to recognize and demand school quality improvement''.

In the same study income, approximated by ownership of a tin roofed house was found to have a positive and significant association with educational demand. Indications are that wealthier parents can afford to send their children to primary schools. In addition to this, it is also reported that large families may derive less utility from sending an additional child to school, because having one more child in school may be less important if some are already enrolled. Some children may be required to help with the family farm or business while others are sent to schools.

A similar study by Wood Hall and Psacharopoulos (1995) showed gender difference to be linked with other important determinants of demand such as attitudes and values pertaining to costs and benefits of sending boy and girl child to school.

The PHRD (1996) report undoubtedly provides the most recent and comprehensive study on demand for schooling in Ethiopia to date. The information for this report came from three different surveys: Two of them, the 1995/96 Household Income, Consumption and Expenditure Survey and Welfare Monitoring Survey, which were carried out by the Central Statistical Authority (CSA) and the third one carried out by the Department of Economics at Addis Ababa University (AAU) in collaboration with the Center for the Study of African Economies in the University of Oxford. The data from the Ethiopia rural household survey, conducted by the AAU and Oxford University concluded that economic constraints such as boys and girls labor for farm activities and school distance represented the most salient impediments to attending school. Similarly the study from the 1995/96 Household Income, Consumption and Expenditure survey and Welfare Monitoring survey showed that the main reasons for the less school participation rate are boys labor, failure, distance to school, inability to afford and low quality of education.

The study based on the 1995/96 Household Income, Consumption and Expenditure survey and Welfare Monitoring survey further revealed that, enrollment ratios are far better in urban areas (56%) than in rural areas (9%) for all school levels. The gender bias in favor of boys is also apparent in nearly all school levels. At national level, the GER was 20% for boys compared to 15% for girls. As can be noted the largest difference is between rural and urban rather than female and male subjects. The bias against girls in the rural areas come out very strongly, especially at primary school levels. The GER for primary school age females in rural areas was only 11% as opposed to 24% for males.

Both GER and NER showed a marked improvement with increase in the educational status of a household head . At national level, for instance, primary GER (for both sexes) increased from 22% in the case of illiterate household heads to 79% for household heads with high school level of education(grades 9-12) and to 102% for household heads with tertiary education. This shows, the big association between the educational status of the household head and school enrollment. The pattern remained the same for urban as well as rural households.

The study also found distance to school to have no effect in urban areas, and that GER stays close to 100% for both sexes. In rural areas, however, as expected, the negative influence of distance was demonstrated by the decline of GER (for both genders) from 29% for households located within 1 km radius from primary schools to only 2% for those located beyond 15 kms. radius from primary school. In the case of secondary education the effect of distance is negative, though its negative impact is high on girls than for boys.

Countrywide (urban + rural) GER increased with increase in per capita expenditure (a proxy for income). For instance GER increased from 13 to 28 when per capita expenditure increased from 0-50 Birr per capita expenditure group to 201 and above Birr expenditure group. This rise of GER with per capita expenditure was observed for both genders at all school levels. This effect becomes more important in rural areas than urban ones. As is generally the case, it is observed that households in higher income groups spent more on education per school-age child and sent a greater proportion of their children to school.

Gender of the household head was also found to be an important determinant. GER and NER revealed that urban-female headed household enrollments at every school level (with the exception of higher education) were higher, for both sexes, relative to that of urban male headed households . In rural areas, female headed households show a slightly higher boy enrollment than male headed households at all educational levels.

In an economy dominated by agriculture, the most important cost of schooling may be the opportunity cost of a student's time. In both household surveys the most important reasons for not attending school were found to be related to opportunity cost (boys for farm activity and girls for other household activity). This means, if children are needed for farm or other household activities, households will less likely send their children to school because of the high opportunity cost involved. However, the two studies concluded that school enrollment in rural Ethiopia is constrained by a combination of demand and supply considerations, and that it is sometimes difficult to separate the two. Therefore it is instructive to consider both the household characteristics which influence demand for education and to examine both anecdotal and survey evidence to provide a full picture of the constraints to education in Ethiopia.

2.4 SHORTCOMINGS OF THE REVIEWED STUDIES

The reviewed studies have some weak points. In some of them important factors that can affect enrollment are not exhaustively considered and some of them are analyzed by simple statistical methods when relatively better methods could be used. For instance the effect of marital status, occupation, health status of the person and etc. were not studied. Even in those studies which used sound statistical techniques the methodology is not discussed and verified very well for its appropriateness for such types of study.

For instance, the PHRD study, which is very recent, is similar as that of our study for its data and descriptive results. But it has not tried to exhaustively consider all relevant variables that can be found from the survey data. Consideration of more interaction variables was not entertained either. The study used the "ENTER" method of analysis when other variable selection methods can be applied. While in our study the forward and backward stepwise methods, which help to easily select factors that can explain the situation as much as all the factors, are used. Indeed, in our case, regression problems like multicollinearity, autocorrelation and outliers are treated by using suitable techniques.

A model developed by any method should be diagnosed for possible problems, otherwise it is difficult to use the results. In this regard the PHRD study had shortcomings, that is, the model was not diagnosed and verified for its appropriateness. If residuals are not analyzed for multicollinearity, autocorrelation and for some influential observations it will not be easy to justify the appropriateness of the model. Classification power, the extent to which the model is correct, should have at least been studied. Besides this other sophisticated statistical techniques were not considered. It is against this backdrop that this study attempts to examine the main determinants of school enrollment.

CHAPTER THREE

3. OBJECTIVES OF THE STUDY, THE DATA AND METHODOLOGY USED

3.1 OBJECTIVES

Education is essential for raising individual productivity, because it gives skills and the basic intellectual tools necessary for further learning (UNESCO, 1993). That is why, prominent on the agenda of social goals is the attainment of priority objectives in the field of education. One among the priority objectives is the attainment of universal primary education.

Despite this prominent agendum, evidence from published statistics indicates that, a strikingly low percentage of school age children in Ethiopia participate in formal education when compared with other low income African countries. According to UNESCO's 1995 statistical year book, the primary GER and most of the other educational indicators show that Ethiopia is far behind most Sub-Sahara African countries. However, only some studies, the majority of which were short of sophistication and completeness in their statistical methods, have been undertaken to identify the main hindrances of enrollment in Ethiopia.

The main objective of this research is therefore to study the main socio-economic and demographic determinants of school enrollment in Ethiopia. Briefly, the specific objectives of this study are as follows.

- To find out the most important factors affecting primary and secondary school enrollment and enrollment in the overall schooling (grades 1 to university);
- To select the key discriminating factors between enrolled and not-enrolled persons of age 7-24 and examine the extent to which the factors discriminate between the two groups;
- To investigate place of residence (urban/rural), inter-regional and gender differences in school enrollment:
- To make recommendations which may be useful for educational planners and policy makers.

3.2 THE DATA

The data for the analysis was procured from the 1995/96 Household Income, Consumption and Expenditure survey and Welfare Monitoring survey carried out by the Central Statistics Authority. The surveys covered the population in the country on a sample basis excepting the non sedentary population in Afar and Somali Regions. A total of 943 Enumeration Areas (EAs), were selected to be covered in all regions by the study. However the survey covered 571 EAs in the rural and 320 EAs in the urban areas

First, for purposes of sampling, the country was divided into 21 rural and 11 urban domains. In rural domains a two stage sample design was used for sample selection, EAs and households being the primary and ultimate sampling units respectively. In the urban case, the urban centers were divided into two groups, i.e. the first group containing ten major urban centers and the second containing the remaining urban centers. In the first group 15 EAs were sampled from each domain, while in the second group 103 EAs were sampled and a three stage stratified sample design was used for sample selection. Here, urban centers, EAs and households served as primary, secondary and ultimate sampling units, in that order.

Sample selection of EAs in all domains was carried out with probability proportional to size (size here is the total number of households according to the 1994 Population And Housing Census Map Work). In each selected EA a fresh list of households was prepared during the field work and 15 were selected from each sampled urban EA and 12 households from each sampled rural EA using systematic sampling technique. Finally, the households in each sample enumeration area were reallocated to Keftegnas /Weredas in proportion to their sizes, size being the total number of households of the of the Keftegnas /Weredas. Overall, a sample covering more than 11,000 households and about 23,000 school age people was collected.

The Household Income, Consumption and Expenditure survey used nine different types of schedules to collect data on characteristics such as age, sex and other demographic variables, consumption of food and drinks, household expenditure and payment on various consumption and non consumption items and household income and expenses . The Welfare Monitoring survey, on the other hand, used seven schedules and data were collected on educational and health status of household members, housing facilities, household assets and anthropometry (CSA, 1997).

Even though the sample size is large , the data have some limitations. The number and the type of explanatory variables is constrained by the available information. Factors like, land ownership, type of shift system, indirect costs and quality of education are not included in this study, because the information was not available. Moreover, some of the relevant variables were with missing cases. The problem that could be created due to the complexity of the survey is also worth mentioning. In applying logistic regression to the data from complex surveys, that is, data from surveys involving clustering, stratification , multistage sampling and non-ignorable non-response, consistent estimators are obtained by appropriately weighing the observations. However, even in this case estimates can be used in the usual case. Because, for large samples asymptotic properties of the estimators remain unchanged (Lehtonen and Pahkinen,1995).

3.3 VARIABLES SELECTED FOR THE STUDY

In my endeavor to understand the main factors affecting school enrollment several studies have been reviewed. Most of these studies (see literature review) showed that schooling determinants, especially in developing countries, are of similar nature. Hence this study uses those variables which are important in light of the reviewed literature and the actual situation of school enrollment in Ethiopia

Describing the study variables is very important in order to understand the application of the methodology and to make a proper usage of the results. Hence, the variables are described as follows. The dependent variables are discrete binary type, that is, they assume the value one if the person is enrolled otherwise zero. Similarly the explanatory variables are classified in to four categories; Viz. personal, household, community and regional dummies (which are included to capture differences in infrastructure, culture and other factors).

A. Dependent variables

1. Primary enrollment (P-ENROL):- it is a discrete binary variable which assumes the value 1 if a sampled primary school age person (ages 7-14) is enrolled in grades 1 to 8 or 0 if not.

2. Secondary enrollment (S-ENROL):- it is a discrete binary variable which assumes the value 1 if a sampled secondary school age person (ages 15-18) is enrolled in grades 9 to 12 or 0 if not.

3. Enrollment (ENROL) :- it is a discrete binary variable which assumes the value 1 if a sampled school age person (ages 7-24) is enrolled in grades 1 to university or 0 if not .

B. Explanatory variables.

I.. Personal variables

1. Age in completed years (AGE)
2. Sex (SEX) 1=Male 0=Female
3. Age squared (AGE-SQ)
4. Relation of the child to the head and spouse (COMCHLD)

   1=Common child 0=Not common child
5. Health status of the person (HEALTH)

   1=Healthy 0=Not healthy
6. Marital status of the person (MAR_ST)

   1= Married 0= Never Married

II Household Variables

1. Family Size (FAM-SIZE)
2. Health status of the head of the household (HLTH_HD)

   1=Healthy 0=Not healthy

3. Age of the head of the household (AGE_HD)
4. Enrollment status of the head (RGSTRN_HD)

   1=Enrolled 0=Not enrolled

5. Literacy status of the head of the household (EDUC-HD)

   1=Literate 0=Illiterate

6. The natural logarithm of expenditure for schooling (LN_SCHEXP)
7. The natural logarithm of total expenditure (LN_TEXP)
8. Sex of the head of the household (SEX-HEAD)

   1= Male 0=Female

9. Literacy status of spouse (EDUC_SPS)

   0= Illiterate, 1= literate

10 . Proportion of School-age girls (PRP_GIRL)

11 Mothers Education in years of schooling (MOTH-EDU)

12. Father's Education in years of Schooling (FATH-EDU)

13. Returns from education (RET-EDU)

   1* = yes 0=no

- *-*If there is one person who earns salary because he is educated*

14. Income of the household (INCOME)

15. Occupation of head (OCUP_HD) or spouse (OCUP_SPS)

1= Farmer 0= Other

16. Employment status of head (EMPL-HD) or spouse (EMPL-SPS)

1=Employed 0= Own employed

17. Number of school age people(age 7-24) (FS724)

III. Community Variables.

1. Distance to Primary School in KMS (DST_PRI)

2. Distance to Secondary School in KMS (DST_SEC)

3. Location of residence (URB-RUR)

1= Urban 0=Rural

IV . Regional Dummies

The regional dummies assume the value 1 if the person under study is from the respective region and 0 if he/she is not from the region. The variables are: TIGRAY, AFAR, AMHARA, OROMIA, SOMALI, BEN-GUM, GAMBELA, HARARI, DIREDAWA, ADDIS.

There are also interaction variables, which are important to detect the interaction effect of some of the factors. These are:

1. The interaction of distance to primary school and location of residence, i.e.

   Urban- Rural (URB_RUR X DST_PRI) & (URB_RUR X DST_SEC)

2. The interaction of mother's education and location of residence

   (URB-RUR X MOTH-EDU).

3. The interaction of sex and location of residence

   (SEX X URB-RUR)

4. The interaction of sex and distance to school

   (SEX X DST_PRI) & (SEX X DST_SEC)

5. The interaction of health and distance to school

   (HEALTH X DST_PRI) & (HEALTH X DST_SEC)

6. The interaction of education and occupation of head

   (EDUC-HD X OCUPHD)

7. The interaction of mother's education and sex of the person

   (SEX X MOTH-EDU).

3.4 METHODOLOGY

A variety of multivariate statistical techniques can be used to predict a binary dependent variable from a set of independent variables. Two amongst these are the logistic regression and discriminant analysis methods. The discussion of both methods is as follows:

3.4.1. LOGISTIC REGRESSION

In most cases the assumption that a probability model is linear in the independent variable is unrealistic. Further if we correctly specify the model as linear, the statistical properties derived under the linearity assumption will not, in general, hold. The obvious solution to this problem is to specify a non-linear probability model in place of the linear probability model (Aldrich & Nelson , 1984; John & Forrest,1984).

For a binary response Y_j and a quantitative explanatory variable X_{ij}, i=1,2,...k and j=1,2,..,n. Let $P_j = P(X_{ij})$ denote the "success probability'' when X_{ij} takes the value x_{ij}. The problem with the linear probability model is that, $\mathbf{E(Y) = X\beta}$ (where β **is** a vector of parameters to be estimated) is used to approximate a probability value, $P_j= P(Y_j = 1)$ constrained to be from zero to one, while $E(Y_j)$ is not so constrained. One approach to this problem is to transform P_j to eliminate one or both constraints. After doing this, we can assume that the transformed variable $\ln\{P_j / (1-P_j)\}$, which lies in the interval $(-\infty, \infty)$, has a linear form for the logit of this probability , that is,

$$\textbf{logit}(P_j) = \textbf{Ln}\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K$$

this implies,

$$P_j = \frac{\mathrm{Exp}(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K)}{1 + \mathrm{Exp}(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K)}$$

where the parameter β_i refers to the effect of X_{ij} on the log odds that Y_j=1 , controlling for other X's. Further more, Exp(β_i) (odds ratio) is the multiplicative effect on the odds of a unit increase in X_{ij} , at fixed levels of the other X's.

3.4.1.1. STATISTICAL INFERENCE

Statistical inference of the model parameters helps us to judge the significance and magnitude of the effects. To test the significance of the effect of X on the binary response, we set the null hypothesis as **H_o: β=0** (the probability of success is independent of X). Since for large sample size the Maximum likelihood (ML) estimate is normally distributed and the Wald statistic (Z^2) is chi-squared distributed with one degrees of freedom, both test statistics can be used to test the hypothesis under consideration. The Wald statistic, which divides a parameter estimate by its standard error and then squared, takes the right tail chi-squared probability above the observed value as its p-value. Whereas likelihood ratio test compares the maximum likelihood function when β=0 (i.e. when P_j is forced to be identical at all X-values) to the ML function for unrestricted β. But for large sample size both these test statistics have the same behavior (Agresti, 1996; Hosmer and Lemeshow, 1989). Hence this study uses the likelihood ratio test, which is more reliable even , in small samples.

Several model selection procedures exist, no one of which is "best". For instance, a model with several predictors has the potential to suffer from multicollinearity, that is strong correlation among predictors making it seem that no one variable is important when all the others are in the model. Hence in our case the backward elimination procedure, which starts with a complex model and that successively removes variables, is used. At each stage, we eliminate the variable in the model that has the largest p-value, when we test that its coefficients equal zero.

It is also important to compare the various models fit to the data in terms of their predictive power. The logistic regression module in SPSS-WIN (SPSS for windows) gives the option R, which is the correlation between binary Y=(0,1) observations on the response and the predicted probabilities p. In such case , R is a crude index of predictive power (Agresti, 1996). However, R has no the nice properties of multiple correlation coefficient in ordinary least square regression. Nevertheless, R is useful for comparing fits of different models to the same data set.

3.4.1.2. MODEL CHECKING

After building a model it is important to examine the adequacy of the resulting model. Residuals comparing observed and fitted counts are useful for this purpose; i.e. as tools for identifying points for which the model does not fit well, points that exert a strong influence on the coefficient estimates, and variables that are highly related to each other.

Fitted logistic regression models provide predicted probabilities that Y=1. At each grouping of the explanatory variable one can multiply the predicted probability by the total number of subjects to obtain a fitted count. The test of the null hypothesis that the model hold compares the fitted and observed counts using a Pearson Chi-Squire (X^2)or likelihood ratio test statistic (G^2). For a fixed number of groups, when most fitted counts equal about five, X^2 and G^2 have approximate chi-squared distributions. The degrees of freedom equals the number of sample logits (i.e. the number of grouping explanatory sample logits) minus the number of model parameters (Agresti, 1996; Fineberg,1997).

Where G^2 is given by

$$\mathbf{G^2 = -2\ Ln\ \{ LLC / LLS\}}$$

LLC = maximum likelihood of constant model

LLS = maximum likelihood of saturated model

A constant model is a model which only contains the constant term, while a saturated or full model is a model containing those variables selected by the stepwise procedure.

As usual, large X^2 or G^2 values provide evidence of lack of fit. When the fit is poor residuals and other diagnostic measures describe the influence of individual observations on the model fit and highlight reasons for the inadequacy . Among these are the residuals (used to detect multicollinearity and other inadequacies), the leverage (used for detecting observations that have a large impact on the predicted values), Cook's distance (a measure of the influence of a case on the residuals), and DfBeta (a measure of the change in the logistic coefficient when a case is deleted from the model) (Draper and Smith,1996). However the logistic regression module in SPSS-WIN does not give further analysis of this results. Hence the values of these diagnostics should be saved and analyzed further using their normal probability plots, plots with each other and the explanatory variables. Still after doing these multicollinearity diagnostics are not that much easy for interpretation. So to avoid this problem a multiple regression model is constructed using the same dependent and independent variables, to search for the VIF(variance inflation factors) and the condition indices. Because multicollinearty has nothing to do with the form of the dependent variable, as it is the correlation between the explanatory variables.

3.4.2. DISCRIMINANT ANALYSIS

The problem that is addressed with discriminant analysis is, how well it is possible to separate two or more groups of individuals given measurements for these individuals on several variables (Manly, 1986). Furthermore discriminant analysis is important for the identification of important discriminating factors of two or more groups.

As mentioned above, our objectives in this particular method of analysis are to examine whether all the variables together permit a greater discrimination among enrolled and not-enrolled children, to see the effectiveness of the discriminating equation in classifying children as enrolled and not-enrolled and to select the most important discriminating variables. To satisfy these objectives, a discriminant analysis was performed using enrollment status of a child as a criterion variable and the factors, which affect school enrollment , as predictor variables. For this purpose among the different options in SPSS-WIN the Wilks' Lambda Minimization stepwise procedure is used. The reason why we are using the stepwise variable selection procedure is because it combines the features of forward selection and backward elimination methods. And that of our interest in the Wilks' procedure is twofold. One is , if the groups are not too close, the criterion gives the most important subset of discriminants, and the second is, it does not require lengthy calculations.

The logic behind Wilks' stepwise procedure is as follows. It starts by calculating a Wilks' Lambda and associated F-to enter for each variable. Then, at the first step, the variable with the largest F-to enter which is greater than the minimum tolerance criterion is chosen, on successive steps one variable is chosen that maximizes the F-value after adjusting for variables already chosen. The stepwise procedure continues until no further significant gain in discrimination can be achieved by the addition of more variables. This step by step procedure is quite important since a variable that is chosen at an early step may, at a later step, be removed because of relationship between it and other variables already chosen (Hawkins (1982); Affifi and Azen (1979)).

There are several features that an " optimal " discrimination function should possess. Firstly, it should result in few misclassifications, that is, the probabilities of misclassification should be small; secondly it may be that one group has a greater likelihood of occurrence than another and hence" prior probabilities of occurrence" should be taken into account; and lastly it should, whenever possible , account for costs associated with misclassifications (Jhonson & Wicherin, 1992) . This study has considered the first two features strictly, but the third case is abandoned as estimating the cost of mis-classifying a child has complications .

To evaluate the discrimination power of the canonical function, the total sample size is divided into two, that is, analysis sample and holdout sample. Using the analysis sample the canonical discriminant function is developed and the first "optimal" feature is checked using the holdout sample. After looking for the probabilities of classification, further evaluation of the discriminant function is done by computing the maximum and proportional chances. In addition to this, to set the prior probabilities for classification we need the priors statement. One option of the prior probabilities in the SPSS-WIN is the sample proportions, that is , the proportion of enrolled & not-enrolled children in the sample. This study uses this option as the prior probability of occurrence.

It is known that the discrimination procedures are based on the assumption of equal within group covariance matrix for both groups. Moreover, tests of significance require the assumption that with in groups the data follow multivariate normal distributions . But , though it may not be simple to establish the significance of results, discriminant analysis is not very sensitive to violations of these assumptions, unless the violations are extreme (Harris, 1975 ; Manly, 1986). Nevertheless, the Statistical package, which has been used for analysis, gives the above two tests by default. Results and interpretations of the analyzed data follow in the next chapter.

CHAPTER FOUR

4. RESULTS OF DATA ANALYSES

In this chapter we shall be concerned with the analysis of the data. It is mentioned in chapter three that the data are collected on a sample basis from two sample surveys. The unit of analysis was a school age person. Even though, the sample size was large in most of the variables the number of missing cases was so large that it was impossible to use all the cases. Hence due to this problem the number of cases used for this analysis are 4555 school age people. Among these 2693 were enrolled and 1862 were not enrolled. To look into the problem widely, the analysis is done firstly, for the primary and secondary school levels separately and next for persons from grade one to university all together. The sample sizes used for primary and secondary schools are 2907 and 889, respectively.

Firstly, a preliminary assessment using a chi square test was performed to determine those factors which are significantly associated with school enrollment. The test showed that most of the variables selected for the study were significantly associated with school enrollment (see Table 4.1A).

Next to the preliminary assessment, further analyses by using logistic regression and discriminant methods is made. In both methods equations which determine the enrollment status of a person from certain predictors are fitted. The differential effect of each factor on gender and location of residence is also investigated by running independent logistic regression as well as discriminant analysis for each factor and their interaction. Results of these are presented below.

Table 4.1A Pearson chi-square Test

Variable	Chi Square Value	DF	P-value
URB_RUR	5803.63911	1	.00000
SEX	69.51615	1	.00000
HEALTHY	27.79978	1	.00000
SEX_HD	350.51348	1	.00000
EDUC-SPS	1777.68501	1	.00000
FATH_EDU	620.04567	13	.00000
MOTH_EDU	443.91820	13	.00000
COMCHLD	15.50422	1	.00008
OCUP_SPS	4565.47598	1	.00000
EDUC_HD	1874.16559	1	.00000
MRST_CHD	1386.41418	1	.00000
ADDIS	1092.91402	1	.00000
TIGRAY	92.51754	1	.00000
AFAR	53.09255	1	.00000
AMHARA	51.06459	1	.00000
OROMIYA	160.59509	1	.00000
SOMALI	7.21749	1	.00722
BEN_GUM	32.68776	1	.00000
GAMBELA	4.90643	1	.02676
HARARI	81.16462	1	.00000
DIREDAWA	57.51039	1	.00000

4.1 LOGISTIC REGRESSION RESULTS

In this method, where the backward elimination technique is used, the variables that make our observed results most "likely" are selected. Estimates of the logistic regression model are presented in Table 4.1B for each schooling level. The table contains the estimated coefficients (under the column heading β) and related statistics.

The logistic models predict school enrollment at each level from a constant and predictor variables listed in the tables. School enrollment, for grades one to university, for instance can be predicted from the variables age of the person, sex of the person, location of residence (urban/rural), his/her relation to the household (to be common child of both the head and the spouse or not), age and health status of the household head, education and occupation of the household head, education and occupation of the spouse, and the number of school age persons in the household. Similarly, proportion of girls, sex of the household head, and school expenditure, are found significantly correlated with primary enrollment in addition to those in the domain. Determinants of secondary schooling, however, are found limited in number. Only location of residence, age of the person, and education of the household head, are found statistically significant. This result is confirmed in the discriminant analysis, except for minor disparity.

The interactions of some of the variables are also found to influence schooling. For instance, interaction of sex and location of residence (URB_RUR) and interaction of sex and distance to school are found significant for the criterion ENROL. The interaction of distance to school and location of residence and the interaction of education and occupation of the household head (to capture the negative effect of being literate but working at home), however are found significantly associated with primary school enrollment. While in the case of secondary school enrollment, the interaction of location of residence and distance to school and the interaction of sex and education of the mother are found significant.

The partial correlation between the dependent variable and each of the independent variables is given under the column R. A positive value of R indicates that as the variable increases in value, so does the likelihood of the event occurring. Obviously if R is negative the opposite is true. As expected age is found to be significantly negatively associated with enrollment at all school levels. This may be an indication of the fact that, those enrolled are very young that, may be those a bit older leave school for different reasons. Further investigation by using a two independent samples t-test showed that the average age of those not enrolled is significantly greater than those enrolled. Normally we expect that the likelihood of going to school to increase up to a certain age limit and to fall there after. It is revealed from the study that, children from the urban centers, children with literate parents, common children of the head and the spouse and male children are found more likely to go to school than the others.

The column Exp(β) is the value with which the odds change when the i^{th} independent variable increases by one unit. In the criterion variable ENROL, education of the spouse and in the case of S-ENROL, education of the head are found to likely increase school enrollment more than all the factors considered. While in P_ENROL, location of residence is with the highest positive influence. Indeed for all school levels location of residence, sex of the person, returns from education and to be common child or not are found with better relative positive influence consecutively. The effect of occupation of the head and the spouse, the interaction of sex with location of residence and distance to school, however, are found to be negative on the probability of school enrollment.

Table 4.1B Variables in the Equation

A. FOR ENROL

Variable	β	S.E.	Wald	df	P-value	R	Exp(β)
AGE	-.2290	.0200	131.527	1	.0000	-.2292	0.7953
AGE_HD	.0173	.0075	5.2715	1	.0217	.0364	1.0174
AMHARA	-.4062	.1818	4.9952	1	.0254	-.0349	.6662
COMCHLD	.5772	.1750	10.8837	1	.0010	.0600	1.7811
DST_SCH	-.0079	.0043	3.3928	1	.0655	-.0238	.9921
EDUC_HD	.2193	.0473	21.4546	1	.0000	.0888	1.2452
FS724	.2859	.0879	10.5748	1	.0011	.0590	1.3310
HEALTH_HD	-.3290	.1509	4.7505	1	.0293	-.0334	.7197
EDUC_SPS	3.0050	.3220	87.0863	1	.0000	.1858	20.1864
OCUP_SPS	-.9620	.2426	15.7206	1	.0001	-.0746	.3821
OCUP_HD	-.4084	.1990	4.2128	1	.0401	-.0300	.6647
SEX	1.1973	.2256	28.1635	1	.0000	.1030	3.3113
TIGRAY	1.0058	.4032	6.2215	1	.0126	.0414	2.7341
URB_RUR	1.9398	.4414	19.3159	1	.0000	.0838	6.9571
INT_1	-1.1998	.2680	20.0450	1	.0000	-.0856	.3012
INT_5	-.1094	.0350	9.7896	1	.0018	-.0562	.8964
Constant	-1.9515	.6647	8.6203	1	.0033		

B. for P_Enrol

Variable	β	S.E.	Wald	df	P-value	R	Exp(β)
AGE	-.1714	.0298	33.0177	1	.0000	-.0740	0.8369
AGE_HD	.0107	.0042	6.4145	1	.0113	.0279	1.0107
COMCHLD	.4506	.1356	11.0422	1	.0009	.0399	1.5693
DST_PRI	-.0731	.0126	33.7784	1	.0000	-.0749	.9295
EDUC_HD	.1068	.0568	3.5393	1	.0599	.0165	1.1127
EDUC_SPS	.7811	.1317	35.1972	1	.0000	.0765	2.1839
OCUP_SPS	-1.3914	.1720	65.4204	1	.0000	-.1058	.2487
OCUP_HD	-1.7096	.4702	13.2201	1	.0003	-.0445	.1809
PRP_GIRL	.4967	.2046	5.8933	1	.0152	.0262	1.6433
RET_EDU	1.3509	.4232	10.1886	1	.0014	.0380	3.8610
SEX	1.0105	.1191	71.9352	1	.0000	.1111	2.7469
SEX_HD	-.7741	.1636	22.3916	1	.0000	-.0600	.4611
TSCH_EXP	.0003	.0001	7.7230	1	.0055	.0318	1.0003
URB_RUR	2.2017	.2295	92.0393	1	.0000	.1260	9.0406
AMHARA	-.7165	.1240	33.3959	1	.0000	-.0744	.4884
GAMBELA	1.6293	.2837	32.9891	1	.0000	.0739	5.1005
OROMIYA	-.5436	.1055	26.5453	1	.0000	-.0658	.5807
SOMALI	-2.4848	.8138	9.3222	1	.0023	-.0359	.0833
TIGRAY	.5488	.1699	10.4371	1	.0012	.0386	1.7311
INT_1	-.9618	.2132	20.3508	1	.0000	-.0569	.3822
INT_3	-.0294	.0088	11.2481	1	.0008	-.0404	.9710
INT_8	1.6842	.4826	12.1778	1	.0005	.0424	5.3882
Constant	-3.4135	.4514	57.1802	1	.0000		

C. for S_ENROL

Variable	β	S.E.	Wald	df	P-value	R	Exp(β)
URB_RUR	3.0808	.7237	18.1202	1	.0000	.1923	21.7753
AGE	-.5151	.2553	4.0686	1	.0437	-.0689	0.6737
EDUC_HD	5.0316	.6728	55.9212	1	.0000	.3516	153.1759
DIREDAWA	-3.9433	1.1823	11.1245	1	.0009	-.1446	.0194
INT_3	1.5087	.4104	13.5131	1	.0002	.1625	4.5207
INT_10	-.1681	.0660	6.4820	1	.0109	-.1014	.8453
Constant	-13.5824	4.5340	8.9741	1	.0027		

Where
INT_1-SEX by URB_RUR
INT_2-DST_PRI by URB_RUR
INT_3-DST_SEC by URB_RUR
INT_4-DST_SEC by SEX
INT_5-DST_PRI by SEX
INT_6-DST_PRI by HEALTHY
INT_7-DST_SEC by HEALTHY
INT_8-EDUC_HD by OCUP_HD
INT_9-MARST_HD by SEX
INT_10-SEX by MOTH_EDU

A model should be assessed and diagnosed for model inadequacies, if it has to be reliable. In this study this is done by using the chi-square goodness-of-fit test and the classification power analyses, shown in Table 4.2 and 4.3, respectively.

Table 4. 2 Goodness of Fit Statistics for the Model

Parameters	ENROL	P-ENROL	S-ENROL
-2 log likelihood (constant model)	2465.2973	5667.9213	436.0859
-2 log likelihood (Saturated model)	1636.9860	3672.983.	113.554
Goodness of fit	2014.3110	4553.949	643.938
Model chi-square	828.3110	1994.938	322.532
degrees of freedom	17	24	6
P-value	.0000	.0000	.0000

A good model is one that results in a high likelihood of the observed results. This translates to a small value for -2LL. Table 4.2 shows that the test for the null hypothesis, that the coefficients for all the terms in the current model, except the constant are zero, to be rejected at P-value less than one percent. This is true in three of the computed models.

Table 4.3 Percentage of Correct Classification

Group	ENROL	P-ENROL	S-ENROL
Not enrolled	72.96	94.95	84.52
Enrolled	83.27	60.61	98.66
Total	78.60	84.02	96.06

Table 4.3 shows the classification powers of the models computed for the criterion variables ENROL, P_ENROL and S_ENROL. . The table shows that the extent of classifying correctly of the three equations are 79, 84 and 96 percent respectively. And these figures are satisfactory to consider the models reliable. Generally it can be inferred from the results that the fit is good.

After assessing the goodness of fit of the test, further model diagnosis is made. As mentioned in chapter three, a multiple regression model is fitted to assess multicollinearity. The collinearity diagnostics, variance inflation factor (VIF) and conditional indices showed that FAMSIZE, FS724 and URB_RUR have significant VIF to cause sever multicollinearity. However, after FAMSIZE, which is highly correlated with FS724, is removed the severity is eliminated. Similar diagnosis for autocorrelation is done using the Durbin-Watson statistic. The value indicates that autocorrelation is not a serious threat to our models (See Table 20 of Annexes).

Residual analysis, that is graphs of residuals and the explanatory variables is also done (See figures in the Annexes). The graphs showed the presence of some outliers (greater than 4 standard deviation. units) which can influence the estimates. However, as the sample size is very large and the number of outliers is insignificant, the outliers are avoided from the analysis. Further residual analysis on the influence measures, DFBETA and Cook's Distance, showed that the estimates are free of influential cases to distort their validity.

4.2 DISCRIMINANT ANALYSIS RESULTS

Discriminant analysis is performed on the data to discriminate between the two groups, enrolled and not-enrolled people. To do this in each criterion variable the sample is divided into analysis and holdout samples. The holdout sample constitutes 17% of the total non-missing cases.

Table 4. 4 Number of cases used for Analysis by Group

Group	ENROL	P_ENROL	S_ENROL
0	1575	936	122
1	2303	1546	640
Total	3878	2482	762

The assumption of two or more groups that are significantly different on any single variable is a prerequisite in performing a discriminant analysis. The univariate F-statistic computed (See Table 5 in the annexes) showed that the two groups are significantly different on about 7 regional dummies and about 15 other socio-economic and demographic factors. From the column Significance, P-values less than 0.01 are considered significant.

The analysis stage of discrimination requires determining and computing the possible number of canonical discriminant functions. Since the analysis involves only two groups, the single

Table 4.6 Canonical Discriminant Functions

criterion	eigen value	Canonical-corr	Wilks' Lambda	Chi-square	df	Sig
ENROL	.5278	.5878	0.654521	1212.638	12	.0000
P-ENROL	.8246	.6723	0.548076	882.467	9	.0000
S-ENROL	3.1023	.8696	0.243765	542.741	5	.0000

possible function computed is presented in Table 4.6. The table contains a chi-square test, significant at P-value less than 0.0001, a canonical correlation coefficient and other characteristics. The chi-square test shows that the two groups have significant difference.

According to Table 4.6 Wilks' lambda values of 0.65, 0.55 and 0.24 are associated with a chi-square (χ^2)=1212.6, 882.5 and 542.7; df 12, 9, and 5, each for ENROL, P_ENROL and S_ENROL, respectively. The other important statistic shown in this table is the canonical correlation coefficient, whose square explains the amount of variation explained by the explanatory variables. It is shown that about 36, 38 and 78 percent of the variation in ENROL, P_ENROL and S_ENROL is explained by the independent variables in the summary table for each of them.

The standardized coefficients which determine the influence of each explanatory variable in determining the canonical discriminant function are presented in table 4.7. However these coefficients are proved to be affected by predictor inter-correlation and they do not reflect common variance. Hence discussion is made using discriminant loadings (structure correlation), which are recently proved to be more valid (Hair et al, 1987).

Table 4.7 Standardized canonical discriminant function coefficients

ENROL	Func 1	**P_ENROL**	Func1	**S_ENROL**	Func1
URB_RUR	.52447	URB_RUR	.39141	COMCHLD	.16317
AGE	-.68750	SEX	.09191	FS724	-.33819
AGE_HD	.09545	AGE_HD	.11254	AGE_HD	.13133
EDUC_HD	.39844	FATH_EDU	.49439	EDUC_HD	.93978
DST_SEC	-.08084	AMHARA	-.15084	DST_SEC	-.16544
FATH_EDU	.18754	OROMIYA	-.08709		
FS724	.24210	COMCHLD	.15123		
PRP_GIRL	-.10216	OCUP_SPS	-.46515		
TIGRAY	.08471	EDUC_HD	-.15462		
SOMALI	-.06868				
COMCHLD	.11279				
OCUP_SPS	-.29058				

The most interesting part of discriminant loadings is that they are ordered by size of their correlation with the canonical discriminant function. Table 4.8 shows that URB_RUR, OCUP_SPS, FAMSIZE, FS724, FATH_EDU, EDUC_HD, PRP_GIRL, EDUC_SPS and AGE for ENROL, URB_RUR, OCUP_SPS, FAMSIZE, FS724, PRP_GIRL, FATH_EDU, EDUC_HD and MOTH_EDU for P_ENROL and EDUC_HD, URB_RUR, FS724, FATH_EDU, , PRP_GIRL, and OCUP_HD for S_ENROL to have the largest influence in the probability of going to school according to their order. Similarly those with the absolute value of their weights between 3.0 and 1.0 have mild influence, while those with the absolute value of their weights less than 1.0 have the least contributions. As can be seen in three of the criterion variables those with the highest influence are similar, though their order of influence differs from school level to school level.

Table 4.8 Structure Matrix: Pooled Within-Groups Correlation Between Discriminating Variables And Canonical Discriminant Functions (Variables Ordered By Size Of Correlation Within Function)

ENROL	Func 1	P_ENROL	Func1	S_ENROL	Func1
URB_RUR	.61577	URB_RUR	.85228	EDUC_HD	.87905
OCUP_SPS	-.58542	OCUP_SPS	-.85228	URB_RUR	.49863
FS724	-.50272	FS724	-.67284	FS724	-.45550
FATH_EDU	.48540	PRP_GIRL	-.57419	FATH_EDU	.37341
EDUC_HD	.47169	FATH_EDU	.57038	PRP_GIRL	-.34863
PRP_GIRL	-.46865	EDUC_HD	.38619	OCUP_HD	-.32030
AGE	-.38146	MOTH_EDU	.36240	MOTH_EDU	.24737
EDUC_SPS	.36850	ADDIS	.28681	EDUC_SPS	.18795
MOTH_EDU	.28947	AGE_HD	.14883	AMHARA	-.12116
DST_SEC	-.26012	TOT_EXP	.13524	AFAR	-.08141
ADDIS	.14885	DST_PRI	-.11131	ADDIS	.07902
MRST_CHD	-.11436	AMHARA	-.10124	MRST_CHD	-.06978
DST_PRI	-.09540	DIREDAWA	.08183	DIREDAWA	.06733
TOT_EXP	.09394	COMCHLD	.07142	SEX	-.06650
COMCHLD	.09274	BEN_GUM	-.05618	BEN_GUM	-.06389
TIGRAY	.08070	OROMIYA	-.04321	COMCHLD	.05537
GAMBELA	-.07983	AGE_SQ	.04287	TOT_EXP	.04692
AFAR	-.07935	AGE	.04115	HEALTHY	.04283
AMHARA	-.07198	HARARI	.04106	TIGRAY	.04131
RET_EDU	-.06727	SOMALI	-.02076	DST_SEC	-.03937
SOMALI	-.05705	GAMBELA	-.02057	HARARI	-.03897
SEX	-.05700	AFAR	-.01464	AGE_HD	.03267
HEALTH_HD	.05608	SEX	.01200	OROMIYA	.03094
BEN_GUM	-.05351	MRST_CHD	-.01127	GAMBELA	-.02587
OROMIYA	.03916	HEALTHY	.01040	AGE	-.01919
DIREDAWA	.02992	TIGRAY	.00357		
AGE_HD	-.02829				
HARARI	.02465				
HEALTHY	.01859				
AGE_SQ	-.01326				

The length of disparity between the groups measured in standard deviation units, that is the number of standard deviation units each group is away from the average of both groups (Hand, 1981), is depicted in Table 4.9. The highest disparity is observed in S-ENROL followed by P_ENROL.

Table 4.9 Canonical Discriminant Functions Evaluated At Group Means (Group Centroids)

Group	ENROL	P_ENROL	S_ENROL
0	-.94453	-1.42094	-3.71799
1	.55844	.57951	0.83012

From the analysis of the preceding section we have seen that the two groups are differentiated from one another on the basis of the variables selected in the study. In that case the group where a person comes was predetermined. But now we want to classify each person to the groups on the basis of the computed functions. For this purpose primarily the prior probability of the person to go to each group should be estimated. Sample proportion, which is unbiased estimate of population proportion, is, therefore used as the prior probability estimator (See Table 4.10).

Table 4.10 Prior probabilities

Group	ENROL	P_ENROL	S_ENROL
0	.37156	.28969	0.18252
1	.62844	.71031	0.81748

Evaluation of the accuracy of a discriminant function mandatorily requires the development of a classification matrix. However the classification method, linear discriminant function (LDF) is best for classification if the assumption of multivariate normality and equal covariance matrices hold true. As a matter of fact our sample size is very large that the law of large numbers will help to assume normality. Moreover, it is indicated in chapter three that discriminant analysis methods are robust for these assumptions if the sample size is large. Hence we can assume that both assumptions are satisfied. For confirmation see the normal probability plot (NPP) and Box's M test in the Annexes.

Table 4.11 Classification Function Coefficients (Fisher's Linear Discriminant Functions)

A. FOR ENROL

ENROL	0	1
URB_RUR	41.7368741	43.7891054
AGE	.9188949	.6294288
AGE_HD	.4230104	.4375925
EDUC_HD	11.0204714	13.3104533
DST_SEC	.1153442	.1063993
FATH_EDU	.2778622	.3542979
FS724	4.2022397	4.4142981
PRP_GIRL	16.5264810	15.8342574
TIGRAY	-3.6310638	-2.8957219
SOMALI	4.4537069	2.4951742
COMCHLD	5.6941539	6.1746027
OCUP_SPS	22.6302191	21.5094311
(Constant)	-48.2594644	-48.2152705

B. FOR P_ENROL

P_ENROL	0	1
URB_RUR	12.9648445	15.1813285
SEX	1.7481407	2.1160407
AGE_HD	.4823344	.5064619
FATH_EDU	.2675162	.5502355
AMHARA	3.6030613	2.8061577
OROMIYA	2.2567450	1.8867024
COMCHLD	6.2661111	7.1381007
OCUP_HD	17.7495539	15.1154889
EDUC_HD	2.9123193	2.4553093
(Constant)	-25.4437991	-27.5599602

C. for S_ENROL

S_ENROL	0	1
COMCHLD	6.5211975	8.5945696
FS724	1.8033027	.5630848
AGE_HD	.5719973	.6340638
EDUC_HD	10.0773981	31.4065142
DST_SEC	.0056535	-.0825125
(Constant)	-21.2116370	-34.8924415

Finally summary classification tables are computed on the analysis as well as the hold out sample from the LDF. These are presented in Table 4.12 and Table 4.13 for the analysis sample and holdout sample respectively. The rows in these tables represent actual group membership, while the columns represent predicted group membership. The entries in the principal diagonal indicate the number of students correctly classified. It is shown in Table 4.13 that the power of classifying a hold out case correctly for ENROL, P_ENROL and S_ENROL are 73, 82 and 76 percent respectively. and similarly that of the analysis sample are 74, 83 and 72 respectively.

Table 4.12 Classification Results for cases selected for Use in the Analysis -

Criterion	Actual Group	Cases	Predicted group membership 0	1
ENROL	Group 0	1575	50.4%	49.6%
	Group 1	2303	9.2%	90.8%
P-ENROL	Group 0	93	67.6%	32.4%
	Group 1	1546	8.2%	91.8%
S-ENROL	Group 0	122	85.2%	14.8%
	Group 1	640	30.0%	70.0%

Table 4.13 Classification Results for cases not selected for Use in the analysis -

Criterion	Actual Group	Cases	Predicted group membership 0	1
ENROL	Group 0	287	53.3%	46.7%
	Group 1	390	12.8%	87.2%
P-ENROL	Group 0	153	69.3%	30.7%
	Group 1	262	10.7%	89.3%
S-ENROL	Group 0	21	81.0%	19.0%
	Group 1	106	24.5%	75.5%

As we can see from the above tables the classification power of the functions are high. Nonetheless, to evaluate the effectiveness of the models completely, we can again compare these correctly classified percentages to the maximum chance criterion (C_{max}) and proportional

Table 4.14 Calculation of Chance Criteria

	ENROL	P_ENROL	S_ENROL
C_{max}	0.59	0.62	0.84
C_{pro}	0.52	0.53	0.73

chance criterion (C_{pro}). Maximum chance criterion is a method of determining the chance classification based on the sample size of the largest group. While the proportional chance criterion is the sum of squares of C_{max} for each group (Mathewos, 1996). Both chance statistics are given in Table 4.14.

Since C_{max} is greater than C_{pro} the maximum chance criterion is the criterion to outperform. The percentage correctly classified for each criterion (0.73, 0.82 and 0.76) exceeds the chance of correct classification for C_{max} criterion substantially. So we again conclude that the dscriminant functions are valid.

The results used so far used all the variables under study for discrimination and classification purposes. As can be seen from Table 4.8 some of the variables are associated with very small coefficients and then their contribution to discrimination would be very small so that selecting the most important or screening the redundant variables for discrimination and classification is of considerable importance. Next an attempt is made to select the "best" subset of variables that classified students to their correct group as good as the whole set of variables.

Table 4.15 Summary Table

FOR ENROL

step	Variables in	Wilks'Lambda	significance	LABEL
1	URB_RUR	.83323	.0000	URBAN/RURAL
2	AGE	.72320	.0000	AGE OF THE PERSON
3	EDUC_HD	.67923	.0000	EDUCATION OF HEAD
4	OCUP_HD	.67301	.0000	OCCUPATION OF SPOUSE
5	FATH_EDU	.66848	.0000	FATHER'S EDUCATION
6	COMCHLD	.66442	.0000	COMMON CHILD?
7	FS724	.66173	.0000	NUMBER OF SCHOOL AGE PEOPLE
8	AGE_HD	.65941	.0000	AGE OF THE HEAD
9	TIGRAY	.65788	.0000	TIGRAY REGION
10	SOMALI	.65653	.0000	SOMALI REGION
11	DST_SEC	.65551	.0000	DISTANCE TO SECONDARY SCHOOL
12	PRP_GIRL	.65452	.0000	PROPORTION OF GIRLS

The summary table (Table 4.15) is the result of the Wilks' stepwise procedure, which contains those variables that classified students into their correct group as good as the whole set of variables. The minimum F-to-enter and F-to-remove are 3.84 and 2.71 respectively. A variable is included in the analysis when it results in the largest F-to-enter which is greater than the minimum F-to-enter. In the end each variable enters and those that result in the largest F-to-remove which is greater than the minimum F-to-remove are selected.

As shown in Table 4.15 the most important determinants of schooling at all levels of schooling are : location of residence, age of the person, education of the head, occupation of the spouse, father's education, to be common child of the head and the spouse, number of school age children, age of the head, distance to school and the proportion of girls. However, in primary schooling, only sex of the child was found out of the domain, that is from those found in all levels of schooling.

B. FOR P_ENROL

Step	Variables in	Wilks'Lambda	significance	LABEL
1	OCUP_HD	.6254	.0000	OCCUPATION OF HEAD
2	FATH_EDU	.5867	.0000	FATHER'S EDUCATION
3	URB_RUR	.6942	.0000	URBAN/RURAL
4	COMCHLD	.5614	.0000	COMMON CHILD?
5	AGE_HD	.55771	.0000	AGE OF HEAD
6	AMHARA	.5423	.0000	AMHARA REGION
7	EDUC_HD	.5162	.0000	EDUCATION OF HEAD
8	SEX	.4972	.0000	SEX OF THE PERSON
9	OROMIYA	.4808	.0000	OROMIA REGION

C. for S_ENROL

Step	Variables in	WilksLambda	Sig.	LABEL
1	EDUC_HD	.29436	.0000	EDUCATION OF THE HEAD
2	FS724	.25664	.0000	NUMBER OF SCHOOL AGE PEOPLE
3	DST_SEC	.25138	.0000	DISTANCE TO SECONDARY SCHOOL
4	COMCHLD	.24679	.0000	COMMON CHILD?
5	AGE_HD	.24377	.0000	AGE OF HEAD

In general, the preceding results of data analysis demonstrate the usefulness of applying multivariate techniques in the analysis of inter-correlated variables. The two methods have resulted in almost the same type of explanatory variables, except for the fact that the order of importance of the variables has a small difference. The study on its way has also confirmed the statement by Johnson and Wichern (1992), which says," when the number of binary type explanatory variables is more than continuos type, the logistic regression approach is better than the discriminant methods; while if the continuos variables satisfy the conditions of normality and constant variation the LDF is better". This is shown by the classification power of the two models, where that of the logistic regression is more than the discriminant method. This does not mean, however, that the discriminant results shown above are invalid. Since the assumption of normality and equal covariance matrices is satisfied we can say that the discriminant results are also equally important.

CHAPTER FIVE

5. DISCUSSION, CONCLUSION AND RECOMMENDATION

In chapter four the results of the analyses together with some discussions were documented. This chapter is devoted to discussion, conclusions and recommendations of the significant factors associated with school enrollment.

As shown in section 4.1, the determinants of the probability of going to school are, age and sex of the person, location of residence, relation of the person to the household head and spouse, age and health status of the head, education and occupation of the head and the spouse, distance to school, returns from education and the number of school age people in the household. Some disparity is observed on the determinants of primary enrollment and secondary enrollment. For instance, total school expenditure and proportion of girls in the household besides to the above factors were found determinants of primary school enrollment, while secondary enrollment was found to be a function of location of residence, age of the person, education of the head, the interaction of distance to secondary and location of residence and the interaction of mother's education and sex of the person only.

Parental education, measured as the literacy status of the parents and in number of school years, influences school enrollment positively. However, the magnitude of the effect varied from male to female. For instance the effect of education of the spouse, that is sound in rural females, is greater for females than for males. Contrary to this, parental education measured in years of schooling had more influence on urban school enrollment. It was found to increase school enrollment probability of urban areas by about 21 percent. This could be attributed to the fact that urban living parents, who are more likely to go to school, have more chance to appreciate the positive returns of investing in education. Hence the government should in some way alleviate the awareness about education of parents. Non formal education, like distance education and adult education could be taken as a means of educating parents.

Occupation of parents had also a significant influence on the probability of school enrollment. Farmers (or in general sense self employed parents) are found to be less likely to send their children to school. The reason could be the need for boys/girls labor is high on farmers than on those government employees, who have more chance to appreciate the returns from education. Of course the negative effect of being self employed, especially in literate parents, is reinforced in rural males. There is no doubt that this problem will be solved with the promising on going situation, the creation of non farming jobs in the rural areas of Ethiopia.

The relation between school enrollment and age of the head was found positive. To say the obvious older parents feel sending their children to school as an investment. Because they feel that they will be helped by their children's income later. Indeed this awareness may not be affluent.

The deterring effect of distance was confirmed by the results of the analysis. For instance, an additional kilometer of distance to secondary school is found to decrease school enrollment by 4 percent. The effect of distance was high in rural areas, especially in rural females. This could be for the reason that distance to school is associated with high cost. That is, when the school is far from the households' residence, the student will be forced to stay out of his/her village the whole school season. This however, has big direct and indirect cost implications. It will require the household to expend additional expenses like, for transport, for meals and mainly the time spent in traveling. For females, especially those at puberty stage, the worry of the parents can have a significant negative effect on their enrollment. However, in urban areas mostly, it is the rich household that demands distant schools in search of quality education from private schools. Hence, if the deterring effect of distance on school enrollment is to be controlled the government should emphasize on constructing more schools in rural areas of Ethiopia. It will be recommendable if the government encourages investment in education from the private sector.

Parents bias in favor of their common children was also observed in this study. This is highly observed in urban areas, especially on their male common children. However, this is not as big as sex bias. Boys enrollment probability is found about three times that of girls. Therefore, to consolidate the increasing trend of school participation in the last five years, the government should give special emphasis on promoting females enrollment, as they are tomorrow's mothers.

Similarly, school participation rates appeared to vary by region and location of residence. As expected school enrollment rates in urban areas are found more than rural areas. But the magnitude varies from region to region. For instance the chance of going to school in Tigray region seems bigger than the Southern Nations , Nationalities and Peoples Region (SNNPR) . Contrary to this the likely of going to school in SNNPR is better than in Amhara. This could be because of the school accessibility, awareness and other social values.

The presence of another school age person in the household was found to increase the probability of school enrollment. This could be attributed to many things. However, in our country especially in the rural areas, double shift systems have not been available. That means parents have no chance to alternatively send their children to school, as the demand for children labor is very high in the rural areas. The effect is indeed aggravated during the main agricultural seasons. Therefore more should have to be done to alleviate the problem. Introducing double shift systems and adjusting the school schedule, when necessary, to less tight seasons could be one way to alleviate the problem.

Girls labor is needed for both house work and agriculture. This makes a person with more sisters more likely to go to school. Because the more the number of girls are the more the person gets substituting labor force.

The interaction of some of the variables had also a significant influence on the probability of school enrollment. The interaction of sex and location of residence, inserted in the equation to capture the reinforced effect of living in rural areas on females, is found to be negatively associated with enrollment. That means, the negative effect of living in rural areas is high on females. Distance to school has also similar negative effect. It is known that rural residing is associated with long school distances. That is why, the deterrent effect of living in rural areas is higher in females.

If the return's from education are harvestable, it is likely that the chances of going to school increases. The interaction term in the equation confirmed this fact. Educated heads who are government employees were found to favor the enrollment of their children. Similar result is found on parents education, measured in years of schooling. This shows, educating parents as a strategy, can increase school participation rate.

Major differences are not observed in the results of logistic regression and discriminant analysis. This shows that both methods are reliable to predict a binary response from a set of related explanatory variables. However, logistic regression seems to outperform discriminant analysis.

The classification power, which was used for model reliability assessment, of logistic regression was found bigger than that of discriminant analysis. Moreover, discriminant analysis does not allow further model diagnosis. Hence, it can be concluded that, both methods may be taken as reliable methods of predicting a binary random variable from a number of predictors.

REFERENCES

Affifi, A. and Azen, S.P. (1979). Statistical Analysis: A Computer Oriented Approach. New York, Academic Press, Inc.

Agresti, A. (1996). An Introduction to Categorical Data Analysis. New York, John Wiley.

Aldrich, H. and Nelson, D. (1984). Linear Probability, Logit and Probit Models. SAGE Publications, Newbury Park.

Birdsall, N. (1987). Demand for Primary Schooling in Rural Mali. Should User Fees be Increased? World Bank.

Chernicovsky, D. (1985). Socio Economic and Demographic Aspects of School Enrollment and Attendance in Rural Botswana. Economic Development and Cultural Change, Vol 6, pp. 303-336.

CSA (1997). Household Income, Consumption and Expenditure Survey Report. Addis Ababa, Ethiopia.

Destafano, J. and Wood Hall (1993) The Demand for primary Schooling in Rural Ethiopia. A Research Study, Addis Ababa.: USAID.

Destafano, J. and Wilder (1992). Ethiopian Education Sector Review. USAID.

Draper, N.R. and Smith, H. (1996) Applied Regression Analysis. John Wiley & Sons. New York.

Fasil G.Kiros (1990). Implementing Educational Policies in Ethiopia. The World Bank, Washington, D.C.

Fineberg, S.E. (1977). The Analysis of Cross-Classified Categorical Data. The MIT Press, Cambridge.

Gertler, P. and Glewwe, P. (1989). The Willingness to Pay for Education in Developing Countries. Washington, D.C., U.S.A.: The World Bank.

Hair, J.F. , Anderson, R.E and Tatham, R.L. (1987).Multivariate Data Analysis. Second Edition.

Hand, D.J. (1981). Discrimination and Classification. New York: John Wiley & Sons.

Harris, R.J. (1975). A Primer of Multivariate Statistics. New York, Academic Press.

Hawkins, D.M. (1982).Topics In Applied Multivariate Analysis. Cambridge, London: Cambridge University Press.

Hosmer, W. and Lemeshow, S. (1989). Applied Logistic Regression. New York, Jhon Wiley.

John H.A & Forrest D.N (1984). Linear Probability, Logit, and Probit Models. SAGE Publications, Inc. U.S.A.

Johnson, R.A. and Wicherin, D.W. (1992). Applied Multivariate Statistical Analysis. Prentice Hall, Inc. Third Edition, U.S.A.

Lehtonen, R. andd Pahkinen, J.E (1995). Practical Methods for Design and Analysis of Complex Surveys. John Wiley & Sons, England.

Maglad, N.E.A. (1994). School Supply, Family Background and gender Specific School Enrollment and Attainment in Sudan. Eastern Africa Social Science research Review, Vol. 10, No. 2, pp. 1-20.

Manly, F.J. (1986). Multivariate Statistical Methods a Primer. Chapman & Hall, London.

Mathewos Tamiru (1996). A Study of Factors Affecting Academic Achievement of Students at Bole Senior Secondary School. Unpublished Master's Thesis, Addis Ababa University.

MOE (1997). Education Statistics. Addis Ababa, Ethiopia.

PHRD (1996). Household Demand for Schooling. Addis Ababa, Ethiopia.

Singh, R.D. (1992). Under investment, Low Economic Returns to Education, and the Schooling of Rural Children: Some Evidence from Brazil. Economic Development and Cultural Change, Vol. 40, No. 3, pp. 645 - 664.

Tan, J.; Lee, K. and Mingat, A. (1984). User Charges for Education. The Ability and Willingness to Pay Malawi. World Bank Staff Paper No. 661, Washington, D.C.

Tekeste Negash (1990). The Crisis of Ethiopian Education. Some Implications for Nation Building. Sweden, Uppsala.

The World Bank (1988). Education in Sub-Saharan Africa: Policies for Adjustment, Revitalization & Expansion. Washington DC.

USAID (1993). Demand for Schooling in Rural Ethiopia. Addis Ababa, Ethiopia.

UNESCO (1993). Trends and Projections of Enrollment by Level of Education, by Age and by Sex. Paris, Division of Statistics.

UNESCO (1995). Trends and Projections of Enrollment by Level of Education, by Age and by Sex. Paris, Division of Statistics.

Wood Hall, M. and Psacharopoulos, G. (1995). Education for Development. An Analysis of Investment Choice, World Bank Publication, Oxford University Press, New York.

ANNEXES

Table 1. DETERMINANTS OF MALE ENROLLMENT

Variable	β	S.E.	Wald	df	P-value	R	Exp(β)
AGE	-.1979	.0255	60.1830	1	.0000	-.2131	.8205
AGE_HD	.0264	.0094	7.8418	1	.0051	.0675	1.0268
EDUC_SPS	2.2776	.4031	31.9296	1	.0000	.1528	9.7535
DST_PRI	-.1059	.0333	10.1188	1	.0015	-.0796	.8996
OCUP_HD	-.6529	.2737	5.6893	1	.0171	-.0537	.5205
FATH_EDU	.1512	.0310	23.7073	1	.0000	.1302	1.1632
COMCHLD	.6327	.2377	7.0883	1	.0078	.0630	1.8828
OCUP_SPS	-1.5335	.2343	42.8351	1	.0000	-.1785	.2158
Constant	-.4905	.6786	.5224	1	.4698		

Overall CLASSIFICATION 74.89%

Table 2. DETERMINANTS OF FEMALE ENROLLMENT

Variable	β	S.E.	Wald	df	P-value	R	Exp(β)
URB_RUR	2.0985	.3366	38.8613	1	.0000	.1760	8.1539
AGE	-.2429	.0302	64.7111	1	.0000	-.2296	.7843
AGE_HD	.0251	.0113	4.9588	1	.0260	.0499	1.0254
EDUC_SPS	3.1266	.5199	36.1630	1	.0000	.1695	22.7961
COMCHLD	.4999	.2421	4.2629	1	.0390	.0436	1.6485
OCUP_SPS	-.7732	.3204	5.8222	1	.0158	-.0567	.4615
EDUC_HD	.4286	.1434	8.9314	1	.0028	.0763	1.5352
Constant	-2.3475	.8189	8.2176	1	.0041		

Overall CLASSIFICATION 80.97%

Table 3. DETERMINANTS OF URBAN ENROLLMENT

Variable	β	S.E.	Wald	df	P-value	R	Exp(β)
AGE	-.3534	.0271	169.6311	1	.0000	-.3750	.7023
EDUC_SPS	2.5693	.4438	33.5165	1	.0000	.1626	13.0566
HEALTH_HD	-.5411	.2190	6.1037	1	.0135	-.0587	.5821
FATH_EDU	.1908	.0318	35.9230	1	.0000	.1687	1.2103
MOTH_EDU	-.0554	.0228	5.9048	1	.0151	-.0572	.9461
COMCHLD	.7365	.2264	10.5835	1	.0011	.0849	2.0887
OCUP_SPS	-.6477	.3141	4.2514	1	.0392	-.0435	.5233
OCUP_HD	-.5640	.2173	6.7366	1	.0094	-.0630	.5689
Constant	2.8330	.5713	24.5947	1	.0000		

Overall classification **83.81%**

Table 4. DETERMINANTS OF RURAL ENROLLMENT

Variable	β	S.E.	Wald	df	P-value	R	Exp(β)
SEX	.7193	.1982	13.1646	1	.0003	.1153	2.0529
AGE_HD	.0388	.0107	13.0648	1	.0003	.1148	1.0396
EDUC_SPS	2.8977	.5443	28.3384	1	.0000	.1771	18.1330
DST_SEC	-.0102	.0047	4.7049	1	.0301	-.0568	.9898
TSCHEXP	.0010	.0003	13.0045	1	.0003	.1145	1.0010
FATH_EDU	.0952	.0404	5.5530	1	.0184	.0651	1.0999
OCUP_SPS	-2.0818	.4804	18.7804	1	.0000	-.1414	.1247
Constant	-4.0566	.8866	20.9335	1	.0000		

Overall classification **76.70%**

Table 5. DETERMINANTS OF ENROLLMENT FOR URBAN MALE

Variable	β	S.E.	Wald	df	Sig	R	Exp(β)
AGE	-.3621	.0396	83.5906	1	.0000	-.3802	.6962
EDUC_SPS	2.2458	.6160	13.2922	1	.0003	.1415	9.4481
FATH_EDU	.2142	.0446	23.1018	1	.0000	.1934	1.2389
COMCHLD	.9008	.3346	7.2465	1	.0071	.0964	2.4616
OCUP_HD	-.7709	.3116	6.1198	1	.0134	-.0854	.4626
Constant	2.6472	.7808	11.4956	1	.0007		

Overall classification 85.28%

Table 6. DETERMINANTS OF ENROLLMENT FOR RURAL MALE

Variable	β	S.E.	Wald	df	Sig	R	Exp(β)
AGE_HD	.0362	.0133	7.3915	1	.0066	.1012	1.0369
EDUC_SPS	3.3243	.6203	28.7210	1	.0000	.2252	27.7790
DST_PRI	-.1087	.0370	8.6056	1	.0034	-.1120	.8970
TSCHEXP	.0009	.0004	6.2318	1	.0125	.0896	1.0009
OCUP_SPS	-2.3346	.7848	8.8496	1	.0029	-.1140	.0969
Constant	-2.9033	1.1537	6.3333	1	.0118		

Overall classification 72.92%

Table 7. DETERMINANTS OF ENROLLMENT FOR URBAN FEMALE

Variable	β	S.E.	Wald	df	Sig	R	Exp(β)
AGE	-.3577	.0382	87.5559	1	.0000	-.3692	.6993
EDUC_SPS	3.1873	.6577	23.4876	1	.0000	.1850	24.2234
HEALTHHD	-1.0767	.2957	13.2609	1	.0003	-.1339	.3407
DST_SEC	-.0397	.0162	5.9827	1	.0144	-.0797	.9611
FATH_EDU	.1324	.0383	11.9723	1	.0005	.1261	1.1415
MOTH_EDU	-.0723	.0316	5.2385	1	.0221	-.0718	.9303
COMCHLD	.7834	.3074	6.4949	1	.0108	.0846	2.1888
Constant	2.7414	.8207	11.1569	1	.0008		

Overall classification 82.86%

Table 8. DETERMINANTS OF ENROLLMENT FOR RURAL FEMALE

Variable	β	S.E.	Wald	df	Sig	R	Exp(β)
AGE_HD	.0378	.0171	4.9007	1	.0268	.0977	1.0386
EDUC_SPS	2.9979	1.0370	8.3568	1	.0038	.1446	20.0431
TSCHEXP	.0008	.0004	4.1180	1	.0424	.0835	1.0008
OCUP_SPS	-2.2480	.5995	14.0636	1	.0002	-.1992	.1056
constant	-3.7576	1.3951	7.2545	1	.0071		

Overall classification 81.46%

Table 9. SUMMARY TABLE FOR URBAN

Step	variables in	Wilks'Lambda	Sig.	Label
1	AGE	.81965	.0000	AGE
2	EDUC_SPS	.75117	.0000	Education
3	FATH_EDU	.73816	.0000	father's education
4	COMCHLD	.72601	.0000	common child?
5	HEALTHHD	.72270	.0000	Healthy?

Percent of "grouped" cases correctly classified: 80.27%

Percent of "grouped" cases correctly classified: 77.46%

Table 10. SUMMARY TABLE FOR RURAL

Step	Variables in	Wilks'Lambda	Sig.	Label
1	EDUC_SPS	.92926	.0000	Education
2	OCUP_SPS	.88461	.0000	occupation of head
3	SEX	.86938	.0000	SEX
4	AGE_HD	.85447	.0000	AGE
5	TSCHEXP	.84014	.0000	total school expenditure
6	DST_SEC	.83328	.0000	Distance to secondary school
7	FATH_EDU	.82760	.0000	father's education

Percent of "grouped" cases correctly classified: 83.58%

Percent of "grouped" cases correctly classified: 82.29%

Table 11. SUMMARY TABLE FOR URBAN MALE

Step	variables in	Wilks'Lambda	Sig.	Label
1	AGE	.78277	.0000	AGE
2	EDUC_SPS	.73182	.0000	Literacy status
3	COMCHLD	.71779	.0000	common child?
4	FATH_EDU	.70549	.0000	father's education
5	OCUP_HD	.69513	.0000	occupation of head

Percent of "grouped" cases correctly classified: 82.99%

Percent of "grouped" cases correctly classified: 82.42%

Table 12. SUMMARY TABLE FOR RURAL MALE

Step	variables in	Wilks' Lambda	Sig.	Label
1	EDUC_SPS	.89971	.0000	Literacy status
2	OCUP_SPS	.85349	.0000	occupation of head
3	DST_PRI	.84011	.0000	Distance of primary school
4	TSCHEXP	.82761	.0000	total school expenditure
5	AGE_HD	.81574	.0000	AGE

Percent of "grouped" cases correctly classified: 81.77%

Percent of "grouped" cases correctly classified: 81.06%

Table 13. SUMMARY TABLE FOR URBAN FEMALE

Step	Variables in	Wilks'Lambda	Sig.	Label
1	AGE	.85103	.0000	AGE
2	EDUC_SPS	.76388	.0000	Literacy status
3	DST_SEC	.74554	.0000	Distance to secondary school
4	HEALTHHD	.72751	.0000	Healthy?
5	FATH_EDU	.71755	.0000	father's education
6	COMCHLD	.70649	.0000	common child?
7	MOTH_EDU	.69967	.0000	mother's education

Percent of "grouped" cases correctly classified: 79.54%

Percent of "grouped" cases correctly classified: 75.68%

Table 14. SUMMARY TABLE FOR RURAL FEMALE

Step	Variables in	Wilks'Lambda	Sig.	Label
1	OCUP_HD	.93909	.0001	occupation of head
2	EDUC_SPS	.90576	.0000	Literacy status
3	AGE_HD	.88411	.0000	AGE
4	TSCHEXP	.86470	.0000	total school expenditure

Percent of "grouped" cases correctly classified: 92.63%

Percent of "grouped" cases correctly classified: 92.85%

Table 15. MALE -FEMALE COMPARISON

Variable	male β	male Sig	female β	female Sig
URB_RUR	.6438	.2446	2.1924	.0083
AGE	-.2097	.0000	-.2537	.0000
HEALTHY	.3180	.3290	-.4909	.0824
AGE_HD	.0232	.0178	.0250	.0307
EDUC_SPS	2.3132	.0000	2.6724	.0000
DST_PRI	-.1143	.0027	.0020	.9562
DST_SEC	-.0040	.4863	-.0059	.4307
TOTEXP	-2.9E-05	.5683	-5.3E-05	.3056
TSCHEXP	6.06E-05	.4347	5.49E-05	.4211
OCUP_HD	-.6034	.0407	-.2249	.4295
MARST_HD	-.7269	.3990	-.0721	.9098
FATH_EDU	.1426	.0003	.1478	.0006
FS724	.2088	.0796	.4300	.0015
PRP_GIRL	-.4723	.3796	.1370	.8358
MOTH_EDU	-.0120	.6898	-.0477	.1105
COMCHLD	.6699	.0088	.7154	.0104
OCUP_SPS	-1.1865	.0006	-.6067	.0997
EDUC_HD	-.0026	.9898	.0328	.8822
MRST_CHD	-.3343	.6259	-.4781	.4428
INT_1	.1748	.1338	-.1362	.2755
INT_2	.0097	.6533	-.0125	.4807
Constant	-.1714	.8779	-1.8848	.1223

Overall CLASSIFICATION 75.54% Overall CLASSIFICATION 80.65%

Table 16. Wilks' Lambda and univariate F-ratio for ENROL

Variable	Wilks' Lambda	F	Significance
URB_RUR	.83323	573.8094	.0000
SEX	.99920	2.3039	.1292
AGE	.92867	220.1989	.0000
HEALTHY	.99915	2.4250	.1195
AGE_HD	.99958	1.2109	.2712
EDUC_SPS	.89491	336.6912	.0000
HEALTH_HD	.99996	.1050	.7459
DST_PRI	.99335	19.1989	.0000
DST_SEC	.96552	102.3971	.0000
TOTEXP	.98983	29.4634	.0000
FATH_EDU	.88939	356.5554	.0000
FS724	.88230	382.4581	.0000
RET_EDU	.99919	2.3148	.1283
PRP_GIRL	.89611	332.3695	.0000
AGE_SQ	.99995	.1473	.7012
ADDIS	.99149	24.6099	.0000
TIGRAY	.99657	9.8559	.0017
AFAR	.99696	8.7282	.0032
AMHARA	.99752	7.1253	.0076
OROMIYA	.99998	.0484	.8259
SOMALI	.99828	4.9255	.0265
BEN_GUM	.99995	.1337	.7147
GAMBELA	.99841	4.5724	.0326
HARARI	.99945	1.5640	.2112
DIREDAWA	.99952	1.3902	.2385
MOTH_EDU	.95459	136.3844	.0000
COMCHLD	.99548	13.0155	.0003
OCUP_SPS	.84681	518.6348	.0000
EDUC_HD	.94482	167.4244	.0000
MRST_CHD	.98699	37.7782	.0000

Table 17. Wilks' Lambda and univariate F-ratio for P_ENROL

Variable	Wilks'Lambda	F	Significance
URB_RUR	.62541	881.6441	.0000
SEX	.99988	.1748	.6759
AGE	.99364	9.4212	.0022
HEALTHY	.99974	.3798	.5378
AGE_HD	.98206	26.8843	.0000
DST_PRI	.98649	20.1584	.0000
TOT_EXP	.97575	36.5885	.0000
FATH_EDU	.78848	394.8796	.0000
FS724	.73584	528.4291	.0000
PRP_GIRL	.78327	407.3074	.0000
AGE_SQ	.99139	12.7876	.0004
ADDIS	.94380	87.6513	.0000
TIGRAY	.99626	5.5279	.0188
AFAR	.99569	6.3683	.0117
AMHARA	.99162	12.4414	.0004
OROMIYA	.99846	2.2660	.1325
SOMALI	.99582	6.1798	.0130
BEN_GUM	.99794	3.0404	.0814
GAMBELA	.99519	7.1163	.0077
HARARI	.99993	.1095	.7408
DIREDAWA	.99044	14.2156	.0002
MOTH_EDU	.91299	140.2849	.0000
COMCHLD	.99581	6.1914	.0129
OCUP_SPS	.62541	881.6441	.0000
EDUC_HD	.89049	181.0200	.0000
MRST_CHD	.99764	3.4766	.0624

Table 18. Wilks' Lambda and univariate F-ratio for S_ENROL

Variable	Wilks'Lambda	F	Significance
ADDIS	.95865	16.6923	.0001
TIGRAY	.99471	2.0573	.1523
AFAR	.98846	4.5195	.0341
AMHARA	.96051	15.9112	.0001
OROMIYA	.99984	.0607	.8055
BEN_GUM	.98846	4.5195	.0341
GAMBELA	.97685	9.1700	.0026
HARARI	.98980	3.9893	.0465
DIREDAWA	.99389	2.3802	.1237
COMCHLD	.99058	3.6804	.0558
OCUP_SPS	.70892	158.8984	.0000
EDUC_HD	.83213	78.0688	.0000
MRST_CHD	.91845	34.3638	.0000
PRP_GIRL	.65992	199.4369	.0000
FS724	.60839	249.1039	.0000
URB_RUR	.54876	318.2218	.0000
SEX	.99421	2.2542	.1341
AGE	.99449	2.1457	.1438
HEALTHY	.98495	5.9123	.0155
AGE_HD	.99670	1.2811	.2584
EDUC_SPS	.29436	927.7334	.0000
DST_SEC	.99521	1.8609	.1733
TOT_EXP	.98743	4.9270	.0270
FATH_EDU	.63039	226.9087	.0000
MOTH_EDU	.92763	30.1935	.0000

TABLE 19. FOR ENROL WITH OUT REGIONAL DUMMIES

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
AGE	.2265	.0199	130.1954	1	.0000	.2280	1.3173
AGE_HD	.0214	.0074	8.3949	1	.0038	.0509	1.0217
COMCHLD	.5944	.1735	11.7432	1	.0006	.0629	1.8120
EDUC_HD	.2128	.0469	20.5860	1	.0000	.0868	1.2371
FAMSIZE	-.1309	.0692	3.5797	1	.0585	-.0253	.8773
FS724	.2980	.0870	11.7427	1	.0006	.0629	1.3471
HEALTH_HD	-.3283	.1503	4.7732	1	.0289	-.0335	.7202
EDUC_SPS	3.0252	.3175	90.7736	1	.0000	.1898	20.5989
OCUP_SPS	-.9556	.2414	15.6761	1	.0001	-.0745	.3846
OCUP_HD	-.3959	.1981	3.9957	1	.0456	-.0285	.6731
RGSTRN_HD	1.6113	.9911	2.6429	1	.1040	.0161	5.0092
SEX	1.2073	.2225	29.4410	1	.0000	.1055	3.3443
URB_RUR	2.0946	.4317	23.5379	1	.0000	.0935	8.1224
INT_1	-1.2015	.2657	20.4483	1	.0000	-.0865	.3008
INT_5	-.1132	.0338	11.2170	1	.0008	-.0611	.8930
Constant	-2.4301	.6464	14.1330	1	.0002		

Table 20. MULTIPLE REGRESSION

Multiple R .61932

R Square .38356

Adjusted R Square .37803

Standard Error .39269

Analysis of Variance

Source	DF	Sum of Squares	Mean Square
Regression	16	171.17873	10.69867
Residual	1784	275.10666	.15421

F = 69.37829, Signif F = .0000

Variable	β	SE(β)	VIF	T	SigT
URB_RUR	.160054	.066394	12.462	2.411	.0160
SEX	.033702	.019681	1.129	1.712	.0870
AGE	-.033819	.002777	1.140	-12.179	.0000
AGE_HD	.003502	.001038	1.207	3.374	.0008
EDUC_SPS	.283202	.036086	1.508	7.848	.0000
DST_PRI	-.010566	.003641	1.398	-2.902	.0038
DST_SEC	-.001164	6.6758E-04	1.607	-1.744	.0813
OCUP_HD	-.077731	.028382	2.043	-2.739	.0062
FATH_EDU	.024647	.003438	2.475	7.169	.0000
FS724	.048002	.013467	9.697	3.564	.0004
PRP_GIRL	-.100949	.055019	2.688	-1.835	.0667
TIGRAY	.148965	.056726	1.054	2.626	.0087
BEN_GUM	.171892	.086951	1.018	1.977	.0482
COMCHLD	.099141	.026344	1.067	3.763	.0002
OCUP_SPS	-.179782	.040938	4.825	-4.392	.0000
(Constant)	.417230	.090411	4.615		.0000

Durbin-Watson Test = 1.51295

TABLE 21. Test of Equality of Group Covariance Matrices Using Box's M

FOR ENROL

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Group Label	Rank	Log Determinant
0	12	-6.308257
1	12	-14.025159

Pooled within-groups covariance matrix

	12	-9.370002

Box's M	Approximate F	Degrees of freedom	Significance
5127.83580	65.42943	78, 16243555.0	0.127109

B. FOR P_ENROL

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Group Label	Rank	Log Determinant
0	9	-6.227156
1	9	-7.800624

Pooled within-groups covariance matrix

	9	-7.057458

Box's M	Approximate F	Degrees of freedom	Significance
23.64297	9.33670	45, 2367631.7	0.270021

C. FOR S_ENROL

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Group Label	Rank	Log Determinant
0	5	-6.227156
1	5	-7.800624
Pooled within-groups covariance matrix	5	-7.057458

Box's M	Approximate F	Degrees of freedom	Significance
423.64297	9.33670	33, 2367631.7	0.235811

FIG.1

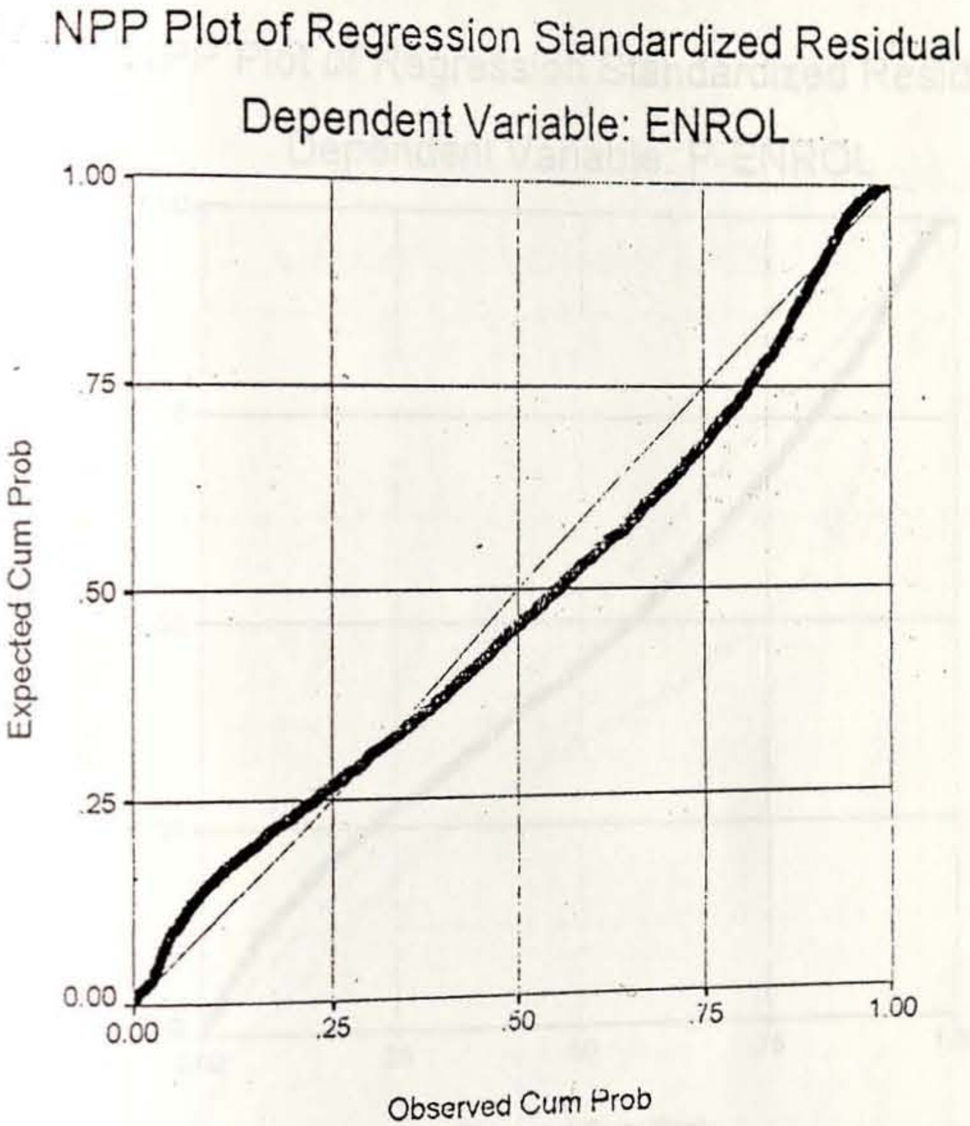

NPP Plot of Regression Standardized Residual
Dependent Variable: ENROL
1.00
.75
.50
.25
0.00
Expected Cum Prob
0.00
.25
.50
.75
1.00
Observed Cum Prob

FIG.2 NPP Plot of Regression Standardized Residual

Dependent Variable: P-ENROL

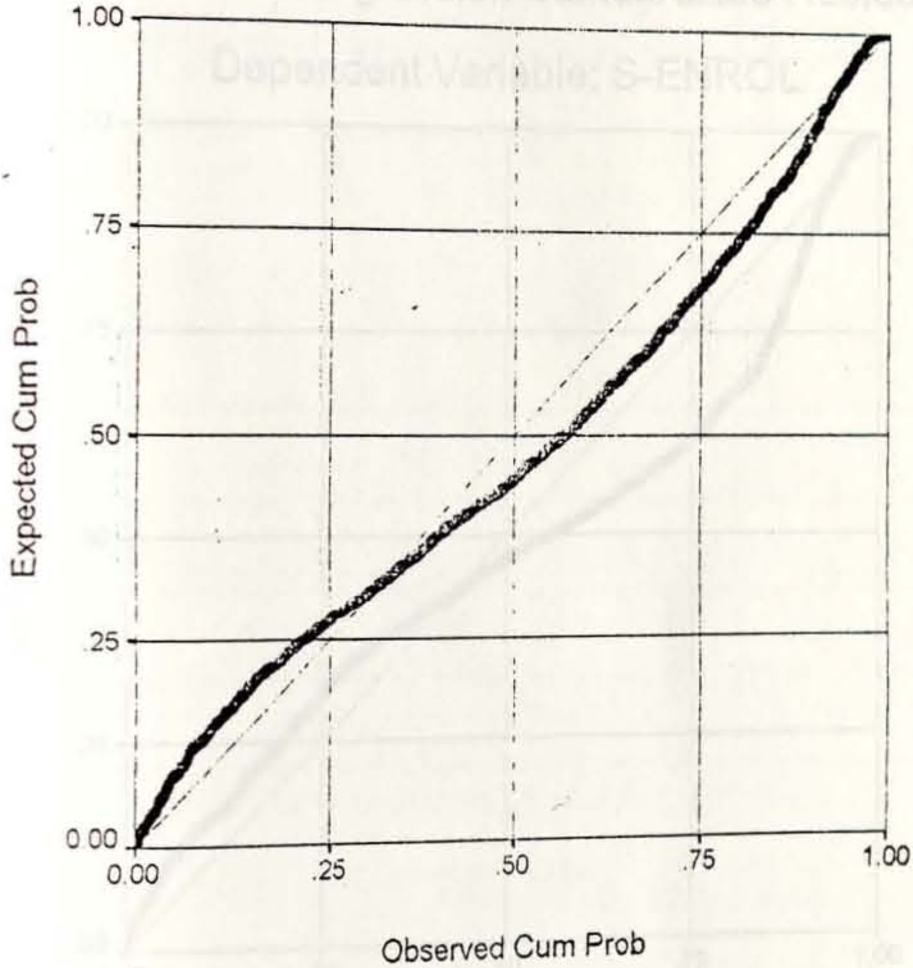


FIG.3

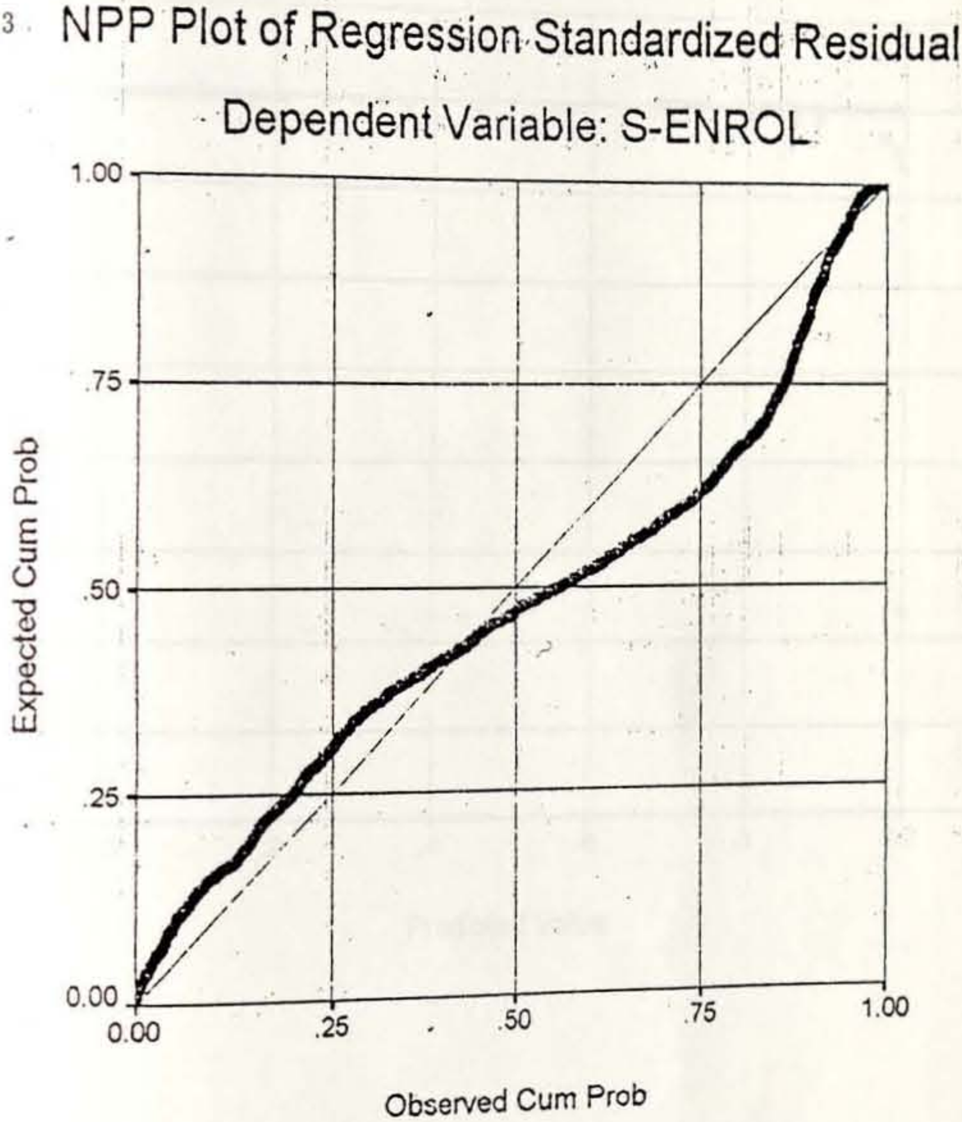
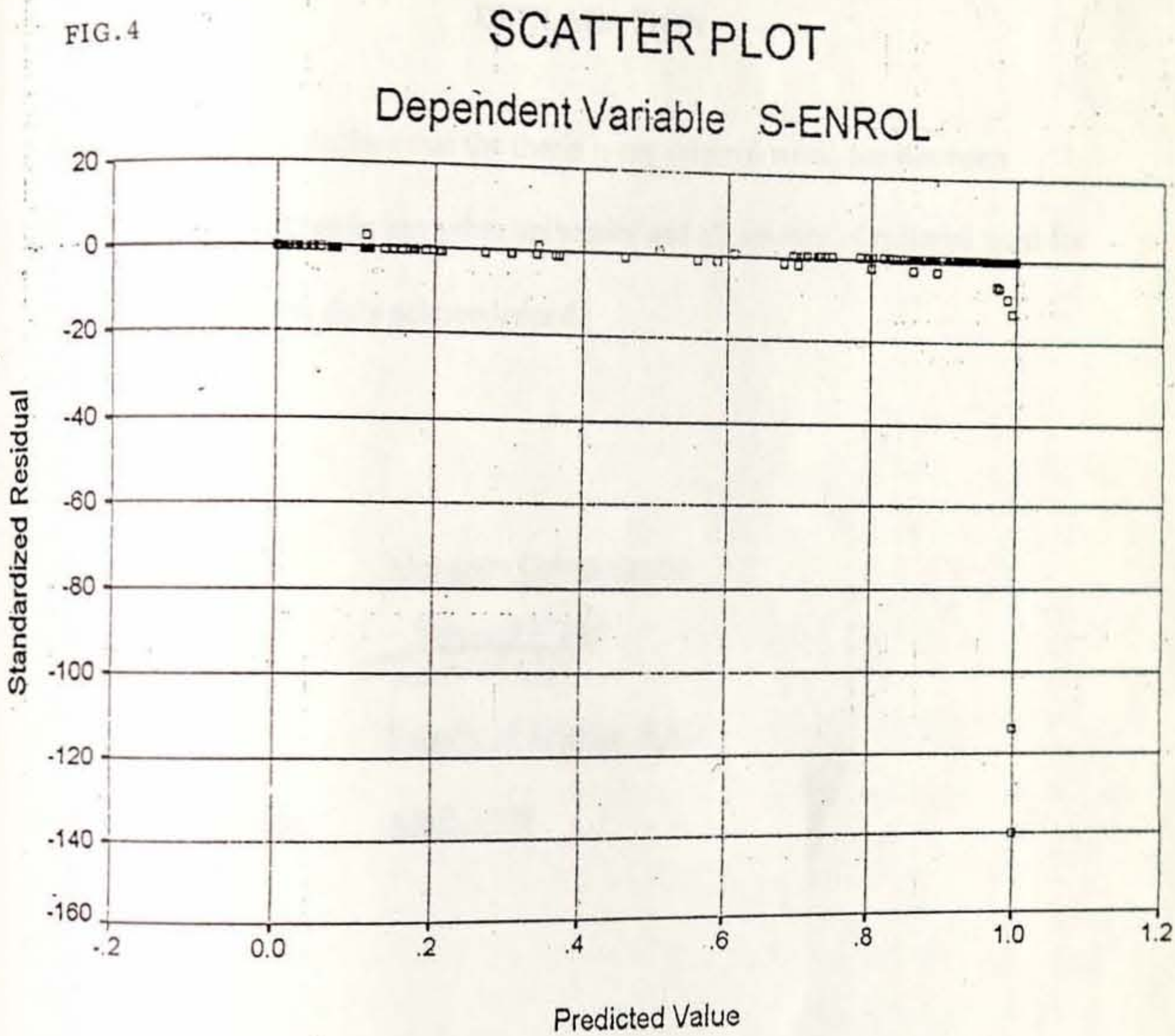

NPP Plot of Regression Standardized Residual
Dependent Variable: S-ENROL
1.00
.75
.50
.25
0.00
Expected Cum Prob
0.00
.25
.50
.75
1.00
Observed Cum Prob

FIG.4



SCATTER PLOT
Dependent Variable S-ENROL
20
0
-20
-40
-60
-80
-100
-120
-140
-160
Standardized Residual
-.2
0.0
.2
.4
.6
.8
1.0
1.2
Predicted Value

DECLARATION

I, the undersigned, declare that the thesis is my original work, has not been presented for a degree in any other university and all sources of material used for the thesis have been duly acknowledged.

Name: Mulugeta Gebreselassie

Signature: 20/04/98

Place: Faculty of Science, AAU

Date of Submission: April, 1998