Copyright Notices

Notice 1

Notice 2

THE IMPACT OF HOUSEHOLD CHARACTERISTICS, SOCIO-CULTURAL FACTORS AND HIV/AIDS ON CHILD WELFARE IN KENYA

by

David Mutie Kimbu

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

to

Department of Econometrics and Business Statistics

Monash University

Australia

MAY, 2011

CONTENTS

Abstract

This thesis investigates socio-economic and health economic issues related to the Kenyan people using the 2003 Kenya Demographic and Health Survey Data. The studies give empirical evidence that helps to understand the impact of socio-economic and health factors, including HIV/AIDS, on child schooling and health status. The thesis examines the factors contributing to an individual acquiring HIV, factors affecting child schooling and the factors having impact on child health in Kenya. Each of these studies employs the econometric tools necessary for reliable estimates. The econometric tools address issues with endogeneity, sample selection and missing data.

The thesis is comprised of five chapters with three major ones. The first Chapter gives an introduction and overview of the thesis. This chapter covers the motivation and objectives of the thesis. It also gives some historical background about Kenya. Chapter 2 examines the factors contributing to an individual acquiring HIV, with a focus on behavioural and personal characteristics, household characteristics and other socio-economic factors.

Chapter 3 deals with children's education outcomes, specifically examining school attendance, school attainment, and rates of grade progression. The study focuses on the impact of individual and household characteristics. It incorporates the effect of socio-cultural factors as well as HIV/AIDS on child schooling. Extensive work is done on relevant econometric and statistical tools designed to address the difficulties associated with the variables used in the analysis. The discussion and conclusion gives some guidance to education policy makers in Kenya and other parts of the world which face similar conditions.

The fourth chapter deals with child health assessment, with a focus on the effect of individual (child) characteristics, household characteristics and other socio-economic and environmental factors on child health. Again due to data and variable problems, suitable econometric tools are used in the analysis of the data. Lastly we finish with Chapter 5 which gives thesis conclusions and future research possibilities.

List of Tables

List of Figures

Declaration

In accordance with Monash University Doctorate Regulation 17 / Doctor of Philosophy and Master of Philosophy (MPhil) regulations the following declarations are made:

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

The core theme of the thesis is the study of welfare of children in Kenya. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the candidate, working within the Department of Econometrics and Business Statistics, Monash University, under the supervision of Professor Brett Inder.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

Signed:

Date:

Acknowledgements

I would like to extend my immense gratitude to my Supervisor, Professor Brett Inder for the professional supervision, inspiration and friendly moral support he has offered to me. My stay at Monash University has been enabled by Professor Brett Inder, who has taken me under his supervision, shown me a great flexibility for me to carry out my research work and studies, notably the Funding provided by the Department of Econometrics and Business Statistics towards my studies. Brett devoted his time beyond the obligations as supervisor to enrich my research until the end.

My utmost thanks go to Katy Cornwell for the technical guidance provided by her in sorting out data and variables used in this thesis. Also, I am grateful to the staff of entire Department for their contribution individually and collectively to my studies in the University. Not forgetting the MEASURE Demographic and Health Surveys (DHS) and Central Bureau of Statistics (CBS, Kenya) for their provision and authorization to use the DHS 2003 Data.

Last but not the least, I am indebted to my family, particularly my wife Mary, my mother Hannah and my children Mercy, Claire and Carla, for the prayers and moral support they have given me throughout.

CHAPTER 1

INTRODUCTION AND THESIS OVERVIEW

1.1 Background and Motivation

In developing countries, policies and strategies for economic development and growth are top priority for economic improvement. The term "economic development" means an increase in economic activity while "economic growth" refers to an increase in levels of wealth and/or income in a particular community or other social unit. Economic development usually leads to economic growth and an increase in the standard of living for many people in the society. Policies towards development ought to improve the economic, political and social well-being of its people. It is common for many countries, however, to pursue policies aimed at short term improvement in wealth and income, abandoning some key factors important for long run economic development and growth, such as investing in children's education and health.

There are several factors that influence growth and economic development. These include human and physical capital accumulation, technology improvement, foreign trade, investment and social attitudes. A number of economic models, from the earlier studies by Harrod-Domar to more recent endogenous growth theory, have sought to provide an integrated framework capturing the essential factors for economic development (See more detail in Chapter 3, section 3.3).

Investment in human capital is an essential factor in economic development. However, capital requires a wide range of inputs in order to accumulate. Becker (1975, 1994) and Ray (1998) elaborate on the factors producing human capital development. These include schooling and expenditure on health care (more detail is given in Chapter 3 section 3.3.). In this thesis, the studies focus on education and health as essential tools for human capital.

Becker (1975) establishes that education is an investment and it adds to our human capital just as other investments add to physical capital. World Bank economists Psacharopoulos (1985; 1994) and Woodhall and Psacharopoulos (1985) indicate that the average return to education (human capital) is higher than that to physical capital in less developed countries (LDCs) but lower in developed countries (DCs). Among human capital investments, they argue that primary education is the most effective for overcoming absolute poverty and reducing income inequality. This is especially true in sub-Saharan Africa, where less than three-quarters of the children of primary school age are enrolled in school.

However, other literature on developing countries suggests higher returns for secondary and higher education since it meets high – level labour requirements of the modern sector rather than establishing literacy and general education as goals for the labour force as a whole (Nafziger, 2006). This is supported by Knight *et al*. (1992) and Knight and Sabot (1990) when they argued that studies supporting higher returns to primary education (given above) are based on methodologically flawed estimates. They established that although average rates of return on primary education were higher than that to secondary education, the marginal rates of return to the cohort entering into the labour market were lower for primary education than for secondary education.

There is evidence of inequality in educational attainment among male and female children, especially in sub-Saharan Africa. This is common in poor households who have to favour educating some children over others due to household income or wealth constraints. Further, the links between parental education, income, and ability to provide a quality education imply educational inequalities are likely to be transmitted from one generation to another. This is usually observed more in the secondary and higher education levels where there is high correlation with parental income and education. There are a number of other factors associated with a child acquiring education. These include individual and household characteristics, socio-economic, cultural, health and environment factors. It is expected that children from a high socioeconomic background are more likely to attend high-cost primary schools, with more public subsidy, better teachers, equipment, and laboratories, and higher school-leaving examination scores, which admit them to the best secondary schools and universities (Nafziger, 2006). It is with this background that the studies in this thesis aim to establish the key factors affecting education in Kenya.

Turning to the health effect on economic development and growth, there is a two-way relationship where development improves the health system, and better health increases productivity, social cohesion and economic welfare in general. Poor nutrition and bad health does contribute not only to physical suffering and mental anguish but also to low labor productivity. A mother malnourished during pregnancy and inadequate food during infancy and early childhood may lead to disease as well as deficiencies in a child's physical and mental development. Future productivity is thereby impaired (Nafziger, 2006).

In another example, life expectancy in less developing countries was observed to have increased steadily between the 1930s and 1994 but due to the HIV/AIDS epidemic in

Africa, life expectancy declined in recent years, from 1994. Due to poor and/or inadequate health facilities, diseases would be one of the major problems responsible for health issues and high mortality rates especially for children. Using 2002 data according to the World Health Organization [WHO] (2003, Annex Table 2.), deaths from diseases, disproportionately from LDCs, include 3.8 million from respiratory infections, 2.8 million from HIV/AIDS, 2.4 million from conditions at birth, 1.8 million from diarrhea diseases, 1.6 million from tuberculosis, 1.3 million from measles, 1.2 million from malaria, and 0.4 million from protein-energy malnutrition and iodine, Vitamin A, or iron deficiency. The report also indicates about 18% of the world's deaths (about 10.5 million) are among children younger than five years old. More than 98% of these child deaths were in LDCs. Kenya is a LDC and it faces the same challenges as other countries.

Although the worldwide child mortality rates fell significantly from 1990 to 2002, Africa's child death rate in 14 countries was observed to have increased. In fact, 19 of the 20 countries with the highest child mortality were in Africa, with the exception being Afghanistan. These child deaths resulted primarily from infectious and parasitic diseases (including HIV/AIDS), conditions at birth, diarrhea diseases, and malaria, with malnutrition contributing to virtually all. Between 1998 and 2003, HIV/AIDS had a huge increase in southern Asia and Southern Africa. The disease was intensified in most regions, with sub-Saharan Africa, Eastern Europe and Central Asia being the worst hit, accounting for about 79% of new infections. The greatest number of people living with HIV were in sub-Saharan Africa as well as in Central Asia (Hussain, 2004). In 2003, all seven countries in Southern Africa had prevalence rates above 17% with Botswana and Swaziland having prevalence above 35%. Adult prevalence in countries in Central and East Africa was in the range of 4 to 13% (Joint United Nations Programme on HIV/AIDS [UNAIDS]; 2004). In

Kenya, the HIV/AIDS pandemic has had a devastating impact on all sectors of the economy, through loss of productivity and labour force (Central Bureau of Statistics [CBS], 2003a).

There is a critical need to carry out this kind of study to build a better understanding of the factors having an impact on child schooling and health status. The study will help education and health policy makers to develop, evaluate and improve existing strategies and policies for the betterment of Kenya's development. Questions like how cultural factors, ethnic diversity, wealth levels and even diseases such as HIV/AIDS impact on economic development can only be answered by carrying out this kind of empirical work. Questions around the underlying factors contributing to individuals acquiring HIV/AIDS in Kenya will also be addressed.

The success of these studies is enhanced by a quality data set. The study uses a national Demographic Survey for Kenya carried out in 2003. The 2003 Kenya Demographic and Health Survey (KDHS) is designed to provide data to monitor the population and health situation in Kenya and to be a follow-up to the 1989, 1993, and 1998 KDHS surveys. The survey information is also intended to provide data to assist policymakers and programme implementers to monitor and evaluate existing programmes and to design new strategies for demographic, social, and health policies in Kenya ((CBS) [Kenya], Ministry of Health [MOH], and ORC Macro., 2004). There is a later survey which was completed in 2008, but this was not available till late in 2010.

1.2 Country Background

This section gives a brief overview of Kenya looking into the historical context, geographical position and economic environment.

1.2.1 Recent History of Kenya

The Portuguese were the first Europeans to explore Kenya, *Vasco da Gama* having visited Mombasa in 1498. By 1730, they had been expelled by the Oman Arabs who by 1839 became powerful due to clove plantations and slave trade under Seyyid Said. With the aim of ending slave trade, the British created a wage-labour system and by the late nineteenth century, the slave trade on the open seas had been completely outlawed by the British. Together with the Germans, the British moved the Oman Arabs out and created trade alliances with influential local leaders in the 1880s. However, the Omani Arab legacy in East Africa is still found today through their numerous descendants found along the coast, typically the wealthiest and most politically influential members of the Kenyan coastal community. The building of the Kenya-Uganda railway passing through the country followed. In the early twentieth century, the interior central highlands were settled by British and other European farmers following the official declaration of Kenyan colonial rule in 1920. This forced the indigenous Kikuyu community to live as itinerant farmers. Africans were excluded from direct political participation until 1944, when the first of them was admitted in the Council. The number doubled to two in 1946, to four in 1948 and to eight in 1951, but these are token politicians, appointed by the colonial governor from local lists.

From October 1952 to December 1959, Kenya was under a state of emergency arising from the Mau Mau rebellion against British rule. The country achieved self-rule in June 1963 and gained independence (*Uhuru*) on December 12, 1963. Exactly one year later, Kenya became a republic. The country has had a stable government and political tranquillity since becoming independent. From the start of its independence until December 2002, the country was ruled by the Kenya African National Union. During the 2002 general elections, the National Alliance of Rainbow Coalition ascended to power through a landslide victory. Internal conflict came in 2005. In 2007 flaws and irregularities in the vote tabulation led to eruption of violence in different parts of the country. This post-election crisis left about 1,300 Kenyans dead and about 500,000 people displaced. These included negative effects on the welfare of children. To resolve the crisis, under the auspices of former UN Secretary General Kofi Annan and the Panel of Eminent African Persons (Benjamin Mkapa of Tanzania and Graca Machel of Mozambique). In February, 2008, President Kibaki and Raila Odinga signed a power-sharing agreement, which provided for the establishment of a prime minister position (to be filled by Odinga) and two deputy prime minister positions. One of the issues addressed was the reform agenda to address underlying causes of the post-election violence. The new constitution was approved in a referendum on August 4, 2010. (Munro, 1975; Kenyatta, 1979; Kanogo, 1987; Wa-Githumo, 1981; Bureau of African Affairs, 2011).

1.2.2 Geography

The Republic of Kenya is a country in Eastern Africa. The country lies between 5 degrees north and 5 degrees south latitude and between 24 and 31 degrees east longitude with a total land area of 580,367 km² (The World Factbook, 2010). It has a population of

38,610,097 (Kenya National Bureau of Statistics [KNBS], 2009). It is almost bisected by the equator and bordered by Ethiopia to the north, Somalia to the east, Tanzania to the south, Uganda to the west, and Sudan to the northwest, with the Indian Ocean running along the southeast border. The country falls into two regions: lowlands, including coastal[1] and lake basin lowlands, and highlands, which extend on both sides of the Great Rift Valley. The Kenyan Highlands comprise one of the most successful agricultural production regions in Africa. Rainfall and temperatures are influenced by altitude and proximity to lakes or the ocean. Under the new constitution, which was passed in August, 2010, the country's administration is no longer provinces and districts. The main administrative unit for the country is a county, 47 counties. At the time of the survey, the country was divided into 8 provinces and 71 districts. These districts were sub-divided into divisions, which were also subdivided into locations and then sub-locations (Counties of Kenya, 2011). The largest cities in Kenya include Nairobi (capital), Mombasa, Kisumu, Nakuru and Eldoret with many other cities and towns spread all over the country (Figure 1) (The World Factbook, 2010; About Kenya, 2004).

1.2.3 Economy

The economy has undergone a structural transformation since 1964. After independence, Kenya promoted rapid economic growth through public investment, encouragement of smallholder agricultural production, and incentives for private (often foreign) industrial investment. GDP grew at an annual average of 6.6% from 1963 to 1973, while Agricultural production grew by 4.7% annually. Between 1974 and 1990, however, Kenya's economic performance declined. Inappropriate agricultural policies, inadequate credit, and poor

[1] The coastline and the port in Mombasa enable the country to trade easily with other countries.

international terms of trade contributed to the decline in agriculture. Lack of export incentives, tight import controls, and foreign exchange controls made the domestic environment for investment even less attractive. From 1991 to 1993, Kenya had its worst economic performance and growth in GDP stagnated and shrank at an annual rate of 3.9%. Inflation reached a record 100% in August 1993. This all led to suspension of donations by bilateral and multilateral donors in 1991. In attempt to revive the economy in 1993, the government eliminated price controls and import licensing, removed foreign exchange controls, privatized a range of publicly owned companies, reduced the number of civil servants, and introduced conservative fiscal and monetary policies. From 1994-96, Kenya's real GDP growth rate averaged just over 4% a year. Economic growth continued to improve and reached 4.3% in 2004 and 5.8% in 2005. In the year 2006, the real growth rate in GDP was estimated to 6.1% (CBS, 2003a; The World Factbook, 2010; About Kenya, 2004).

In 2000, The World Factbook, (2010) estimates about 50% of the population was below the poverty line. Kimalu *et al.* (2002) estimates that the national headcount index increased from 52.32% in 1997 to 56.78% in 2000. Poor growth in the economy contributed to deterioration in the overall welfare of the Kenyan population. Similarly, the economy had been unable to create jobs at a rate to match the rising labour force. The worsening living standard is also shown by rising child mortality rates, increasing rates of illiteracy, and rising unemployment levels (CBS, 2003a; The World Factbook, 2010; About Kenya, 2004).

The HIV/AIDS pandemic has also had a devastating impact on all sectors of the economy, through loss of production and labour force. Against this background, the government of

Kenya in 2003 launched the Economic Recovery Strategy for Wealth and Employment Creation, aimed at restoring economic growth, generating employment opportunities, and reducing poverty levels (CBS, 2003a). The government is convinced that employment creation is the most effective strategy for halting the increasing poverty.

1.2.4 Religion affiliation

Approximately 68% of Kenyans are Christians (35% Protestant, 23% Catholic and 10% Seventh-day Adventist), about 18% are Muslim,[2] 10% are Traditional Religions and 4% Others, which include the Asian community - Hindus, Sikhs, Parsees, and Bahais (The World Factbook, 2010; About Kenya, 2004; African Studies Center [ASC], 2008).

1.2.5 Ethnic groups and languages

There are over 70 distinct ethnic groups in Kenya, ranging in size from about seven million Kikuyu to about 500 El Molo who live on the shore of Lake Turkana. Kenya's ethnic groups divide into three broad linguistic groups: Bantu, Nilotic and Cushite.[3] The largest ethnic group, the Kikuyu, makes up only 22% of the nation's total population, followed by, Luhya (14%), Luo (13%), Kalenjin (12%), Kamba (11%), Kisii (6%), Meru (6%), Other African (15%) and non-African (Asian, European and Arab) (1%). The Kikuyu, who were most actively involved in independence are disproportionately represented in public life,

[2] Over half of Kenya's Muslim minority are of Somali origin. The remainder is largely made up of Galla-speaking peoples and the Swahili-speaking community on the coast, which has maintained uninterrupted contact with Muslims from the Arabian peninsula since the fourteenth century.

[3] The Kikuyu, Meru, Gusii, Embu, Akamba, Luyha (or alternate spelling of Luyia), Swahili and Mijikenka (which in fact is a group of different ethnic groups) constitute the majority of the ***Bantu*** speaking peoples of Kenya. In general, the Bantu are agriculturalist. ***Nilotic*** ethnic groups include the Luo, Masai, Turkana, Samburu, and the Kalenjin and are pastoralist and fishermen. ***Cushitic*** speaking people comprise a small minority of Kenya's population and include: Somali, El Molo, Boran, Burji Dassenich, Gabbra, Orma, Sakuye, Boni, Wata, Yaaka, Daholo, Rendille, and Galla. They are basically nomads and pastoralist in the dry and arid lands of Kenya.

government, business and the professions. The principal non-indigenous ethnic minorities are the Arabs who mostly live in Coast Province, and Asians, who mostly live in Nairobi city (Kurian, 1992; The World Factbook, 2010; About Kenya, 2004; ASC, 2008).

1.3 Objective of the Thesis

The main objective of this thesis is to use a nationally representative Kenyan data set to explore the factors affecting child schooling and health status, to inform policy and strategies aimed at improving long run economic development. We use suitable econometric tools with consideration of specific issues related to the data and variables used in each case.

1.4 Thesis Outline

This thesis consists of three papers, given in three main chapters dealing with different issues. The first chapter examines the impact of household and individual characteristics on the likelihood of an individual adult acquiring the HIV disease. The model includes social factors both measured at the level of the household and the individual. These include household residence, wealth, individual tribe, religion, occupations, level of education, marital status and other behavioural factors.

The second chapter examines child schooling outcomes, specifically school attendance, grade attainment and grade progression rate. Each model examines the impact of household characteristics, socio-cultural factor and HIV/AIDS on these three child schooling outcomes. School attendance assesses a child's current classroom engagement and school attainment highlights the likelihood of children completing primary education. The model

of grade progression rate assesses the ratio of the current grade already completed to the number that the child should have completed given their age, so is a measure of productivity of the education process. The study employs different econometric models with specific tools to estimate models. The results reveal a range of factors that can explain child schooling outcomes.

The third chapter investigates the factors influencing the health of children. These factors include the child's characteristics, household and mother's characteristics. To measure child health, the study uses three indicators: body mass index (*bmi*) for age, height for age and weight for age. Z-scores are obtained to standardise these indicators to have mean zero and variance one. The results indicate several factors that are responsible for child health and survival.

Fig 1. Map of Kenya showing the Provincial Administration

CHAPTER 2

FACTORS CONTRIBUTING TO INDIVIDUAL HIV/AIDS INFECTION IN KENYA

2.1 Introduction

The HIV/AIDS epidemic stands as a threat to global welfare, with its tremendous potential to destroy families' hopes and livelihoods. Kenya is not isolated from this threat and in this chapter we are going to look into the factors which could possibly play a part in an individual being infected with HIV. Literature indicates that behaviour and lifestyle could be contributing to the spread of HIV/AIDS. Illiteracy, cultural practices, traditional beliefs and misconceptions on HIV/AIDS are other contributing factors to an individual's infection with HIV/AIDS (Zozie, 1998). The impact of HIV/AIDS on Kenyan communities includes not only loss of beloved family members, but it also leads to poverty and deterioration in the country's economic outlook (Adams, 2006).

HIV stands for Human Immunodeficiency Virus and is the virus that causes AIDS. In the body, the immune system controls many viruses and other pathogenic diseases and the failure or break down of the system can provide an avenue for these diseases to our body and could lead to death. HIV targets and infects the same immune system cells that are supposed to protect us from illnesses. These cells are a type of white blood cell called CD4 cells (T-cells) and HIV takes over these CD4 cells and turns them into virus factories that produce thousands of viral copies. As the virus grows, it damages or kills CD4 cells thus weakening the immune system (The Well Project, 2010; Hunt, 2010).

The term AIDS stands for Acquired Immune Deficiency Syndrome. AIDS is the most advanced stage of HIV infection which occurs after HIV attacks the immune system. When the immune system loses too many CD4 cells, a person is less able to fight off infection and can develop serious, often deadly, infections. These are normally called opportunistic infections (OIs) because they take advantage of the body's weakened defense mechanism. In fact, when someone dies of AIDS, it is usually the opportunistic infections or other long-term effects of HIV infection that causes death. Therefore, AIDS can be looked as the body's immune-compromised state that can no longer stop OIs from developing and becoming so deadly to someone's life (The Well Project, 2010).

From the above definition, it is clear that a person does not have AIDS at the moment they are infected with HIV, but the effect of HIV with time causes a person to eventually develop AIDS, due to deterioration of the immune system in the body. An individual can be HIV-positive for many years with no signs of disease, or only mild-to-moderate symptoms, but without treatment, HIV will eventually wear down the immune system to the point that they develop more serious OIs (The Well Project, 2010; Hunt, 2010).

2.2 Chapter Literature Review

2.2.1 Brief History of HIV/AIDS

There is no clear evidence of the source of HIV, but scientists suggest that it is likely that HIV first appeared in humans in West Africa near the beginning of the twentieth century as a result of infection by simian immunodeficiency virus (SIV) from chimpanzees. It is most likely that the virus jumped to humans when humans hunted these chimpanzees for meat

and came into contact with their infected blood. Over several years, the virus slowly spread across West Africa and later into other parts of the world including other parts of Africa.

In their research work, Bailes *et al.* (2003) determined that the ancestry of HIV-1 (human immunodeficiency virus-1) had been traced to SIVcpz (simian immunodeficiency virus) infecting chimpanzees (Pan troglodytes)] in west central Africa. However, they are very clear that the origin of SIVcpz itself remains unknown. A more recent study by Keele, *et al.* (2006) revealed the history of the virus, linking the HIV-1 virus to a closely related simian immunodeficiency virus (SIV) discovered in captive chimpanzees of the subspecies *Pan troglodytes troglodytes* originating in South eastern Cameroon where prevalence rates in some communities reached 29 to 35%[4]. There has been several theories about zoonosis[5] which includes the hunter theory (Wolfe, *et al.,* 2004), the oral polio vaccine (OPV) theory (Cohen, 2000; Blancou, *et al.,* 2001; Berry, *et al.,* 2001), the colonialism theory (Chitnis, *et al.*, 2000) and the conspiracy theory (Fears, 2005).

One of the areas where this disease spread initially was the Caribbean. It was when this disease appeared in the homosexual population of the United States that AIDS gained public attention and the virus has existed in the United States since at least the mid- to late 1970s. From 1979-1981 opportunistic infections (OIs), which included rare types of pneumonia, cancer, and other illnesses were being reported by doctors in Los Angeles and New York among a number of male patients who had sex with other men. These were conditions not usually found in people with healthy immune systems (Hunt, 2010; Centers for Disease Control and Prevention [CDC], 2008).

[4] See Jonathan, *et al.* (2006)
[5] When a viral transfer between animals and humans takes place

In 1981, clusters of cases of Kaposi's sarcoma were reported in young patients in San Francisco and New York which was an unusual occurrence in the United States.[6] Since then other OI diseases associated with immuno-compromisation arose in this same population and during 1982, similar immunodeficiencies were found in hemophiliacs, persons who received blood transfusions and intra-venous drug users who shared needles. At this point, it was clear that an infectious agent was involved and this agent was either passed during sexual intercourse or by receiving blood (or blood products) from another person. During this time when these events were occurring in western countries, doctors in Uganda were observing a similar fatal wasting syndrome that they called slim disease (Hunt, 2010; Centers for Disease Control and Prevention [CDC], 2008).

To this date, the world pandemic of HIV/AIDS lives with us and a cure has not been found. Despite major success in treating infected people in western countries, the disease has become the major cause of death in many third world countries in which treatment and control measures reach only a minority of the infected population. Scientists have all along attempted to develop a vaccine but have so far been unsuccessful. With the prevalence of HIV in the developing world, HIV and its complications will be with us for many generations to come. AIDS is now a leading cause of death worldwide for those over 5 years of age (Hunt, 2010; Centers for Disease Control and Prevention [CDC], 2008).

2.2.2 Global HIV/AIDS Statistics

There are about 33 million HIV-infected people in the world of whom around 22.5 million are in sub-Saharan Africa where the adult infection prevalence is about 6%. Approximately

[6] Kaposi's sarcoma was a rare disease that normally occurred in elderly men of Jewish or Mediterranean ancestry.

14,000 new HIV infections occur daily around the world and over 90% of these are in developing countries. One thousand are in children less than 15 years of age. Of adult infections, 40% are in women and 15% in individuals of 15 - 25 years of age. Prenatal infection has resulted in a large number of children being born with HIV. 30-50% of mother-to-child transmissions of HIV results from breast feeding and about a quarter of babies born to HIV-infected mothers are themselves infected.

In recent years, HIV infections have levelled off in the west and the wave of infections threatening to affect western heterosexuals has not materialized. This has not been the case elsewhere, with a huge increase in southern Asia and southern Africa. HIV prevalence has intensified in most regions, with sub-Saharan Africa, Eastern Europe and Central Asia being the worst hit, accounting for about 79% of new infections between 1998 and 2003. The greatest number of people living with HIV were in sub-Saharan Africa as well as in Central Asia (Hussain, 2004). In 2003, in six countries, adult HIV prevalence was below 2%, while in six other countries it was over 20%. In southern Africa all seven countries had prevalence rates above 17% with Botswana and Swaziland having prevalence above 35%. Adult prevalence in countries in Central and East Africa was falling, and in the range of 4 to 13% (UNAIDS, 2004).

With time, the epidemic appears to have stabilized in most regions, although prevalence continues to increase in Eastern Europe and Central Asia and in other parts of Asia due to a high rate of new HIV infections. Sub-Saharan Africa remains the most heavily affected region, accounting for 71% of all new HIV infections in 2008. In 2007, Swaziland had the most severe level of infection in the world with an adult HIV prevalence of 26%. Despite high rates, in 2008 Lesotho’s epidemic seemed to have stabilized, with a prevalence of

23.2%. South Africa continued to be home to the world's largest population of people living with HIV with about 5.7 million in 2007. In Kenya, young women are three times more likely to become infected than their male counterparts with young women between the ages of 15 and 19 being particularly vulnerable to HIV (UNAIDS, 2009)

Despite the high rates in sub-Saharan African, there has been growing evidence of HIV prevention successes in diverse settings with incidence declining in some countries. A drop in HIV incidence was reported among women in Zambia between 2002 and 2007. In Tanzania, national HIV incidence fell between 2004 and 2008. Zimbabwe experienced a steady fall in HIV prevalence since the late 1990s, due to changes in sexual behaviour (Hallett *et al*., 2010). This was partly due to the improved access to treatment where the Antiretroviral therapy coverage rose from 7% in 2003 to 42% in 2008, with especially high coverage achieved in eastern and southern Africa (48%) (World Health Organization [WHO], United Nations Children's Fund [UNCEF] & Joint United Nations Programme on HIV/AIDS [UNAIDS], 2009).

HIV/AIDS is a significant challenge in developed and developing countries alike. But in developing countries, HIV/AIDS threatens to reverse decades of economic development because it attacks people in their most productive years, destroys communities, disrupts food production and places heavy burdens on already weak health services (Australian Government - AusAID, 2009). The associated economic and social implications due to morbidity and mortality affect the education sector, making children more vulnerable and life expectancy continues to decline. Despite increasing management options, adequate control is still out of reach. The disease is causing social and economic havoc around the world, especially in Sub-Saharan Africa. The reasons why the HIV/AIDS pandemic is so

severe in Sub-Saharan Africa is an issue that has not been fully understood. There is need to investigate the effect and extent of HIV/AIDS in this region.

2.2.4 Factors Associated with an Individual Acquiring HIV/AIDS

HIV infection and spread to other humans occurs after some contact with infected blood or blood-associated fluid (semen, vaginal fluid and mother's milk). This is specifically through unprotected sexual intercourse with an infected person (either heterosexual or homosexual), transfusion of infected blood or blood products, transfer from infected mother to her baby during pregnancy, the birth process or through breast–feeding, and also the use of infected needles and instruments without sterilization or sharing of needles and syringes, especially by HIV-infected drug addicts. Despite the awareness and attempts to control and prevent oneself being infected by HIV/AIDS, by nature there are day-to-day challenges and factors which make an individual more likely to acquire the disease. This chapter analyses some of the factors which may contribute to acquiring or spreading HIV to individuals and society.

Lau and Muula (2004) discuss possible contributory factors to HIV/AIDS. One factor is income inequality and lack of social cohesion. Lau and Muula claim that the economic crises associated with unfair international trade continue to contribute to the impoverishment of many southern African countries. This, with many other factors including population relocation, inequality, civil unrest, infrastructure prone to increased mobility and changing beliefs, exposes Sub-Saharan Africa to a high-risk environment for the spread of an infectious disease. The interplay of contemporary behavioural patterns and biological factors facilitate the extensive spread of HIV-1 infection. The least number of HIV infections is caused by transmissions via blood transfusions and infected needles. In

most situations, sexual behaviour is affected by socioeconomic and cultural factors. In Sub-Saharan Africa, the subordinate position of women in society, impoverishment and the decline of social services, and rapid urbanization and modernization have played a significant role. The reason why and how urbanisation would facilitate the spread of HIV in Africa makes some sense. Firstly, the migration of people from rural into urban areas results in the loss of traditional and cultural values and can leads to multiple sexual partnerships. In urban areas, individuals are no longer under the influence of the rural traditional environment and its sanctions and consequently adopt an individualistic lifestyle. Another reason is the high population in urban areas as compared with rural areas, which, for some people, acts as an incentive to become involved in extramarital or multiple partner sexual relations without being discovered by one's partner (Lau and Muula, 2004).

In their study, Bedi, *et al.* (2004) report that HIV/AIDS was first reported in Kenya in 1984, and between 1990 and 1999 the HIV prevalence rate increased from 4.8% to 13.5% where the prevalence of AIDS is considerably higher in urban areas and in adults. Persons aged between 15 and 49 account for around 94% of the total number of HIV-positive persons. Most deaths associated with AIDS occur in the age range 29 to 39 years old.

In previous related work, Sateren *et al.* (2006) have used logistic regression to estimate odds ratios to determine the prevalence and risk factors for HIV – 1 infection among agricultural plantation residents in Kericho, Kenya. The study only concentrates on impact using an ecologically small area and the estimates are likely to be biased due to sample selection. The respondents were volunteers and non-random. In her study, Fortson (2008) uses various methods to estimate the link between wealth and education and HIV status. She employs both nonparametric and parametric models with quadratic functions to

establish the relationship between years and of schooling and HIV and between wealth and HIV. Due to interpretation difficulties, she estimates model with education and wealth included separately. However, by doing that, there was a possibility of these variables capturing the effect of each other and hence giving biased estimates. In addition, there are other relevant factors which could explain the variation of HIV. Wealth could also be an endogenous variable and not correcting for this could give biased estimates.

2.3 Organisation of Survey and Data Processing

The Data was obtained from the 2003 KDHS, a national level population and health surveys. The 2003 KDHS is designed to provide data to monitor the population and health situation in Kenya and to be a follow-up to the 1989, 1993, and 1998 KDHS surveys. There is a later survey which was completed 2008, but was not available until late 2010.

2.3.1 Sample Design

A representative probability sample of almost 10,000 households was selected for the KDHS sample. This sample was constructed to allow for separate estimates for key indicators for each of the eight provinces in Kenya, as well as for urban and rural areas separately. Given the difficulties in travelling and interviewing in the sparsely populated and largely nomadic areas in the North Eastern Province, a smaller proportion of households was selected in this province. Urban areas were over-sampled. As a result of these differing sample proportions, the KDHS sample is not self-weighting at the national level; consequently, the analysis is based on weighted data.

The survey utilised a *two-stage sample design*. The first stage involved selecting sample points (*clusters*) from a national master sample maintained by CBS (NASSEP IV). A total of 400 clusters, 129 urban and 271 rural, were selected from the master frame. The second stage of selection involved the *systematic sampling* of households from a list of all households that had been prepared for NASSEP IV in 2002[7]. All women aged 15-49 years, either usual residents of the households or visitors present in the household on the night before the survey, were eligible to be interviewed. In addition, in every second household selected for the survey, all men age 15-54 years were eligible to be interviewed if they were either permanent residents or visitors present in the household on the night before the survey.

New features of the 2003 KDHS include the collection of information on malaria and the use of mosquito nets, domestic violence, and HIV testing of adults (Central Bureau of Statistics [CBS] [Kenya], Ministry of Health [MOH] [Kenya] & ORC Macro, 2004). The new information included in this survey made it appropriate and suitable for this study.

2.3.2 Questionnaires

Three questionnaires were used in the 2003 Kenyan DHS survey. These include the Household Questionnaire, the Women's Questionnaire and the Men's Questionnaire[8]. The contents of the questionnaires were based on model questionnaires developed by the MEASURE *DHS* programme. In consultation with a broad spectrum of technical institutions, government agencies, and local and international organisations, CBS modified

[7] The household listing was updated in May and June 2003 in 50 selected clusters in the largest cities because of the high rate of change in structures and household occupancy in the urban areas.
[8] These questionnaires were translated from English into Kiswahili and 11 other local languages (Embu, Kalenjin, Kamba, Kikuyu, Kisii, Luhya, Luo, Maasai, Meru, Mijikenda, and Somali).

the DHS model questionnaires to reflect relevant issues in population, family planning, HIV/AIDS, and other health issues in Kenya (CBS, MOH, & ORC Macro, 2004).

2.3.3 HIV testing

In all households selected for the Men's Questionnaire, all women and men who were eligible for the individual interview were asked to voluntarily give a few drops of blood for HIV testing. The protocol for the blood specimen collection and analysis was based on the anonymous linked protocol developed by the DHS programme and approved by ORC Macro's Institutional Review Board, revised and enhanced by KEMRI and CDC. This was further reviewed and approved by the Scientific and Ethical Review Committees of KEMRI and by the Institutional Review Board and Director of CDC in Atlanta, Georgia (CBS, MOH & ORC Macro, 2004).

2.3.4 Response rates

A total of 9,865 households were selected in the sample, of which 8,889 (about 90%) were occupied and therefore eligible for interviews. The unoccupied were either vacant or destroyed structures. Out of these eligible households, only 8,561 were successfully interviewed, yielding a household response rate of 96%. From 8,717 identified and eligible women, the interviews were completed with 8,195 of these women (response rate of about 94%). For the 4,183 eligible men, who were identified in the sub-sample of households selected for the male survey, only 3,578 were successfully interviewed a response rate of 86%.[9] Rural areas had higher response rates compared with urban areas, for both males and

[9] The principal reason for non-response among both eligible men and women was the failure to find individuals despite repeated visits to the household and even sometimes the work place. The substantially lower response rate for men reflects the more frequent and longer absences of men from the household.

females. In total, 8,486 individuals were eligible for HIV testing. From these, 76% of 4,303 eligible women and 70% of eligible men were tested for HIV. The descriptive statistics in Table 2.2 give more details about the HIV prevalence showing in the data (CBS, MOH & ORC Macro, 2004).

2.4 Econometric Methodology

This section gives the variable definitions used in this chapter and the empirical model specification for our analysis. The aim of this chapter is to try and identify the social and demographic factors that are likely to contribute to an individual being infected with the disease, and conversely, the factors likely to protect an individual from getting the disease. The dependent variable measures whether a person is infected, and the explanatory variables are characteristics of the individual that might affect incidence of the disease. Of course the choice of variables is restricted to the data that is available in the DHS data set.

2.4.1 Variables in the Model

We construct a set of independent variables that may contribute to an individual becoming infected with HIV. These include individual and household characteristics (household head, household wealth, parent's marital status, ethnicity, religion, occupation, etc) and demographic and geographical factors.

The model includes a dummy variable for gender of the household head. The Individual's education level was divided into categories: no education, incomplete primary school education, with primary school education, with at least secondary school education and higher education. The model includes a dummy variable for if a person is from an urban

area, dummy variables for the 8 provinces, 13 variables for ethnic groups or tribes[10], 3 for religion, 8 for occupation, permanent and seasonal work, dummy variables for husband or partner staying or living away or at home, and the husband having more than one wife (polygamy). In addition we included a dummy variable for household member gender taking a value of 1 for male and 0 for female, the variable for household wealth which was continuous, constructed using Principal Component Analysis (PCA) (more detail is given in the section 2.4.3). Included also was a dummy variable for male circumcision. Circumcision is a cultural and religious activity practiced by almost all tribes in Kenya[11] and it is today viewed as an unhygienic and traumatizing activity to females, and classified as a high potential for spreading HIV. Research in Boyle (2002) and Hoffman (2002) indicates that apart from the medical consequences, female circumcision has sexual consequences. The complete list and definition of variables is given in the Appendix of this chapter.

2.4.2 Estimation Issues

In our study we will use a probit model to estimate a model for an individual's HIV status. However, there are two main issues with the estimations using this data: household wealth endogeneity and the HIV sample selection. One of the factors that might influence HIV status is a person's income or wealth. In the DHS survey data, wealth is given in the form of household assets from which wealth quintiles are computed; there is no continuous income or wealth measure. Wealth is likely to be endogenous because there are many individual-specific factors that contribute to an individual building up wealth which might

[10] The included tribes are Embu, Kalenjin, Kamba, Kikuyu, Kisii, Luhya, Luo, Masai, Meru, Mijikenda/ Swahili, Somali, Taita/taveta, Turkana, Kuria, and Others (includes native and foreigner tribes in Kenya apart from the name ones)

[11] According to their traditions and culture, males from Luo community are not supposed to be circumcised.

also affect the person's HIV status. Some of these factors are observable and others unobservable. In our model we have controlled for most of these factors but due to the fact that some are not observed, the wealth variable is treated as an endogenous variable.

Relatively little attention has been paid to models with an ordinal endogenous independent variable, including the problem of obtaining the correct standard errors. Simple 2SLS methods exist which account for endogeneity in models where either the dependent or the independent endogenous variable is continuous or dichotomous (Arendt & Holm, 2006; Alvarez & Glasgow, 2000), whereas such procedures are generally not consistent with qualitative endogenous variables. Other combinations of dependent variables including dichotomous, polychotomous or censored require the use of computationally involved methods (see Nelson & Olsen, 1978). The mostly commonly used method is two-step estimation introduced by Amemiya (1978) and Heckman (1978) where the properties of these techniques were explored by Rivers and Voung (1988) and also by Alvarez and Glasgow (2000). Maddala (1983, pp. 242-7) provides corrections for standard errors in two-step methods for six models of different types of variables, but not for the ordinal endogenous variable with more than two categories (See also Hill & Water, 1995; Schotz & Lubell, 1988). Some researchers have ignored this kind of presence of endogeneity in an ordinal variable and estimated equations separately or transformed and truncated the data in order to use the available statistical tools. This will give biased estimates. Some researchers have carried such analysis without correcting the standard errors. Claibourn and Martin (2000) and Nelson and Olsen (1978) are good examples.

To deal with this problem I use principal component analysis to construct a continuous wealth variable, and then use the Rivers Voung procedure to correct for endogeneity of wealth (Rivers & Vuong, 1988).

The second problem concerns non-random selection of the sample for HIV testing. The data was collected from individuals who agreed to be tested, with other individuals refusing to be tested for HIV status, given that they belonged to a selected household. Due to this problem, there is a possible bias in the estimates. A Heckman procedure is used to deal with this non-random sample selection (Heckman, 1978; 1979).

Two methods were used to select the number of principal components to include in the main equation to proxy wealth. Firstly, we used the criterion developed by Kaiser (1960) called the Kaiser Criterion. With this approach, you retain and interpret any component with an eigenvalue greater than 1.00. The basis of this criterion is that each observed variable contributes one unit of variance to the total variance in the data set. Any component with an eigenvalue greater than 1.00 is accounting for a greater amount of variance than had been contributed by one variable. In this case, such a component is therefore accounting for a meaningful amount of variance, and is worthy of being retained (Kaiser, 1960). Another method we use to confirm this component retention was the scree plot. In this method, the eigenvalues associated with each component are plotted. One is required to get the component just before a "break" between the components with relatively large eigenvalues and those with small eigenvalues. Again, the components that appear *before* the break are assumed to be meaningful and are retained for rotation, whereas those appearing *after* the break are assumed to be unimportant and are not retained (Cattell, 1966). Using these two methods, component one was retained as a continuous proxy for

household wealth. This measure is almost certainly measured with error, another reason to allow for potential endogeneity of this wealth variable. This endogeneity is dealt with using standard instrumental variables, with asset variables and other household characteristics (eg. Education level of household head) as instruments.

2.4.3 Principal Component Analysis (PCA)

As noted, the DHS data set used in this chapter does not provide a continuous measure of wealth. Instead, based on household assets in the data, a principal components analysis was used to construct an index of wealth. This data set then allocates each household to one of 5 wealth categories from poorest to richest depending on their value for the wealth index. So one option for capturing wealth effects on HIV status is to include this set of dummy variables capturing the different wealth categories for the household. However, there are two problems with this. First, it discards information about wealth – a continuous wealth index has been replaced with a set of 5 categories. There is a great deal of variation within these categories that is ignored here. Secondly, wealth is almost certainly endogenous and dealing with an endogenous ordinal variable is not easy. Most researchers have avoided this kind of problem, using different set-ups to tackle the problems at hand. Mukhopadhyay *et al.*, (2011) gives a detailed overview on this issue. For these reasons, we have followed the DHS process, but stopped one step from the end. Namely, we use principal components analysis on the asset indicator variables to construct, using the first principal component, a continuous index of household wealth.

The principal component analysis approach was used by Filmer and Pritchett (1999; 2001) and Montgomery *et al*. (2000), among others in constructing the asset index as a proxy for wealth. PCA extracts, from a set of variables, the few orthogonal linear combinations of the

variables that capture the common information most successfully (Shlens, 2005). This condenses our selected household assets into a small set of factors. This is valid if the intercorrelations between our variables are significant but not excessive. Bartlett (1937) came up with the test for sphericity (Bartlett test) which enables us to assess the validity of PCA by assessing if the computed correlation matrix of the included variables is significantly different from the identity matrix (Sigmund & Carlson, 1969; Snedecor & Cochran, 1989). To test for sphericity using the Bartlett's Sphericity Test, consider *p* variables and *n* observations and define

$$\chi^2\left[\frac{\left(p^2-p\right)}{2}\right]=-\left[\frac{(n-1)-(2p+5)}{6}\right]\ell_n\,|R| \tag{2.1}$$

where $|R|$ is the determinant of the matrix (generalized variance) of the sum of products and cross-products from which the intercorrelation matrix *R* is derived, the degrees of freedom for the χ^2 statistic is $\left(p^2-p\right)/2=p(p-1)/2$ which is the number of off-diagonal correlations, and ℓ_n denotes the natural logarithm. The null hypothesis (H_0) is that the given variables are not intercorrelated. If H_0 is not rejected, then variables are not significantly different from a "spherical" set of variables (totally uncorrelated), and there is little point in doing a PCA. As a caution, the problem of the intercorrelations being excessive would mean the presence of multicollinearity and this problem is identified by use of the Kaiser-Meyer Olkin (KMO) statistic which takes values between 0 and 1. A small value indicates that the variables have too little in common to warrant a PCA or factor analysis (Kaiser, 1970). The overall KMO statistic should be at least 0.6. In our variable selection for PCA analysis, we followed Kaiser's criterion, eliminating the

household variables which had KMO less than 0.6 (Kaiser, 1974; Dziuban & Shirkey, 1974).

2.4.4 Sample Selection Problem

A sample selection problem occurs when the observed sample is not a random sample but systematically chosen from the population. From the DHS survey for HIV data, 8,697 individuals aged 15 – 54 years were selected randomly for HIV testing. However, of these individuals only 6,360 (73.12%) agreed to be tested for HIV. Those who refused comprised 13.63% of the sample and the rest were not present for testing or there were technical or other problems with the test. This quite high non-participation rate for the testing brings the issue of sample selection which might cause bias in estimation.

To deal with this problem, we need to specify an equation to model the decision an individual makes about agreeing to be tested. Only if they agree do we observe the test result. This is a classic selection problem, so could be solved with the standard Heckman procedure. Now assuming the two error terms are jointly normally distributed, we can estimate the model by the Heckman two-step procedure (or by Maximum Likelihood). The Heckman two-step procedure involves estimating a probit for the decision to agree to be tested, then using these estimates to construct an Inverse Mills Ratio (IMR), which is included as an additional variable in the main equation for individuals' HIV status.[12] The IMR is given by the probability density of the probit index divided by the cumulative density of that probit index and is used to correct the selection bias in the main equation. In

[12] Basically the variables used in main equation to control for unobservables that are correlated with the selection of the HIV individuals in the estimation sample.

the Stata, we used "heckprob" which employs purely maximum likelihood estimation and does not estimate ρ directly, but atanh ρ (see equation 3.19 in Chapter 3 section 3.5.3).

In the selection model, we included instruments for the individual agreeing to participate in the HIV testing or not. We identified two explanatory variables as instruments for the selection model which were not included in the main equation. The first variable was a dummy for if there was an interruption of the interview by the husband or any other adult male or female from the household. If there was an interruption of the interview by either of the mentioned persons, it indicates the respondent was in fear in making her own decisions and most likely was reluctant to be tested for HIV disease. The second variable was for the religion which was represented by dummy variables for Catholic, Protestants and Seventh Day Adventist variables, Muslim and other or no religion. Since religion can influence individuals' moral and social behaviour, it would be highly likely to influence an individual's chance acquiring of the HIV virus and the decision to undergo HIV testing. An individual associated with any such faith may be in fear of being found to be HIV positive, and as a result, they would not be willing to go for any test. They may fear being seen as having deserted from teaching of their religion.[13] To check at which stage it had significant influence, we included these dummy variables in the HIV equation and selection equation separately and in both equations together. The coefficients showed influence only for the selection equation. On this basis, we took the variables as instruments for the selection equation only.

[13] See Kowalewski, (1990); De Waal, (2006)

2.4.5 Empirical Specification

The structural estimation equation for the individual's HIV status is specified as:

$$HIV_i^* = X_i'\beta_1 + W_i^{*'}\beta_2 + u_{1i} \quad (2.2a)$$

$$HIV_i = \begin{cases} 1 \text{ if } HIV_i^* > 0 \\ 0 \ \ otherwise \end{cases} \quad (2.2b)$$

where HIV_i^* is an unobserved latent variable with X_i as the set of covariates comprising the individual, household and demographic characteristics, W_i^* is the variable measuring household wealth, β is the vector of the parameters we want to estimate and u_1 is the error term such that $u_{1i} \sim N(0,1)$ and independent.

Since W_i^* is endogenous, to correct this, firstly a continuous household wealth variable is constructed using principal component analysis:

$$\left.\begin{array}{l} W_i^A = pc\{X_{Ai}\}, \\ \text{the first principal component } (c_1 = X_{Ai}'\alpha_{1i}) \end{array}\right\} \quad (2.3)$$

where c_1 is the first component extracted, α_{1i} is the regression coefficient (or the weight) for the observed household asset variable X_{Ai}, W_i^A is the variable measuring the household wealth.

Now since the variable W_i^A is known to be endogenous it is likely to contribute to biased estimates. To correct for this problem, we specify the following equation:

$$W_i^* = X_{Wi}'\gamma + u_{2i} \quad (2.4)$$

where in this case W_i^* is the behavioural equation for household wealth. We assure $Cov(u_{1i}, u_{2i}) \neq 0$ and $E(u_{2i} \mid X_{wi}) = 0$. X_{wi} is a $(1xl)$ vector of factors which could contribute to household wealth which includes X_i from equation (2.2a) such that $(k+1 \leq l)$ and part of γ that corresponds to the excluded variables includes at least one non-zero element. To correct for bias due wealth endogeneity, we predicted the wealth residuals from a regression estimation of equation (2.4) and use these residuals as an additional variable in our main equation (2.2). This is the Rivers and Vuong procedure.

To correct for the bias due to sample selection, we used the Heckman procedure where we specify:

$$S_i^* = X_{Si}^{'}\psi + u_{3i}, \text{ and } S_i = \begin{cases} 1 \; if \;\; S_i^* \left(= X_{Si}^{'}\psi + u_{3i}\right) > 0 \\ 0 \;\; otherwise, \quad i = 1, 2, \cdots, n \end{cases} \tag{2.5}$$

where S_i^* is the latent variable and S_i is the probit equation for whether the individual's HIV test outcome is observed. $u_{3i} \sim N(0,1)$. Assume $corr(u_{1i}, u_{3i}) = \rho \neq 0$.[14] We require X_{si} to have at least one variable that is not in the probit equation, or the model will be identified only by functional form yielding coefficients which have no structural interpretations.

The above set of equations could in principle be estimated by maximum likelihood. This method is more efficient but is sensitive to the assumption of multivariate normality of u_{1i}, u_{2i} and u_{3i}. These kind of models are complex and often ML estimation does not converge,

[14] If $\rho = 0$, the standard probit will give consistent estimates.

especially when ρ is far from zero, and because of the existence of local maxima, the result may not be correct even if the procedure converges. This is common in such structural models (Nawata, 1994). Consequently we have opted for the slightly more ad hoc, but hopefully more reliable, approach of augmenting the main equation with the appropriate residuals to give consistent estimates. While there is a small loss in efficiency compared with maximum likelihood, the estimation is much less vulnerable to convergence and stability problems.

Define IMR as

$$\lambda_i\left(\alpha_{u_3}\right)=\frac{\phi\left(\alpha_{u_3}\right)}{1-\Phi\left(\alpha_{u_3}\right)}=\frac{\phi\left(X_{si}^{'}\psi/\sigma_{u_3}\right)}{\Phi\left(X_{si}^{'}\psi/\sigma_{u_3}\right)} \tag{2.6}$$

We can re-write equation (2.2) as

$$HIV_i^{*}=X_i^{'}\beta_1+W_i^{'}\beta_2+\beta_\lambda\lambda_i\left(\alpha_{u_3}\right)+v_i \tag{2.7}$$

where $\beta_\lambda=\rho\sigma_{u_1}$

Since our interest is on the effect of independent variables on HIV status, our aim is to first obtain consistent estimates of β_j. To interpret these estimates we need to then compute the marginal effects (Greene, 2003 pp.782).

To find the marginal effects of the regressors on these response probabilities, we obtain the partial derivatives. Assuming x_{ji} is a continuous variable belonging to x_i (containing w_i) and x_{si}, then from equation (2.6) the effect of a change in x_{ji} on the expected value of HIV_i^{*} for individuals in the population given that HIV_i is observed is:

$$\frac{\partial E\left(HIV_i^* \mid S_i^* > 0\right)}{\partial x_{ji}} = \beta_j + \beta_\lambda \frac{\partial}{\partial \alpha_{u_3}}\left[\frac{\phi\left(\alpha_{u_3}\right)}{1-\Phi\left(\alpha_{u_3}\right)}\right]\frac{\partial \alpha_{u_3}}{\partial x_{ji}}$$

$$= \beta_j + \beta_\lambda \frac{\phi'\left(\alpha_{u_3}\right)\left[1-\Phi\left(\alpha_{u_3}\right)\right]+\left[\phi\left(\alpha_{u_3}\right)\right]^2}{\left[1-\Phi\left(\alpha_{u_3}\right)\right]^2}\left(-\frac{\psi_k}{\sigma_{u_3}}\right) \qquad (2.8)$$

Recall: $\lambda_i\left(\alpha_{u_3}\right) = \frac{\phi\left(\alpha_{u_3}\right)}{1-\Phi\left(\alpha_{u_3}\right)} = \frac{\phi\left(X_{si}'\psi/\sigma_{u_3}\right)}{\Phi\left(X_{si}'\psi/\sigma_{u_3}\right)}$ and since $\phi'\left(\alpha_{u_3}\right) = -\alpha_{u_3}\phi\left(\alpha_{u_3}\right)$ equation (2.8) becomes

$$\frac{\partial E\left(HIV_i^* \mid S_i^* > 0\right)}{\partial x_{ji}} = \beta_j + \beta_\lambda \left\{\frac{-\alpha_{u_3}\phi\left(\alpha_u\right)\left[1-\Phi\left(\alpha_{u_3}\right)\right]}{\left[1-\Phi\left(\alpha_{u_3}\right)\right]^2} + \frac{\left[\phi\left(\alpha_{u_3}\right)\right]^2}{\left[1-\Phi\left(\alpha_{u_3}\right)\right]^2}\right\}\left(-\frac{\psi_j}{\sigma_{u_3}}\right)$$

$$= \beta_j + \beta_\lambda \left\{\frac{-\alpha_{u_3}\lambda_i\left(\alpha_{u_3}\right)\left[1-\Phi\left(\alpha_{u_3}\right)\right]\left[1-\Phi\left(\alpha_{u_3}\right)\right]}{\left[1-\Phi\left(\alpha_{u_3}\right)\right]^2} + \frac{\left[\phi\left(\alpha_{u_3}\right)\right]^2}{\left[1-\Phi\left(\alpha_{u_3}\right)\right]^2}\right\}\left(-\frac{\psi_j}{\sigma_{u_2}}\right)$$

$$= \beta_j + \beta_\lambda \left\{\left[\lambda_i\left(\alpha_{u_3}\right)\right]^2 - \alpha_{u_3}\lambda_i\left(\alpha_{u_3}\right)\right\}\left(-\frac{\psi_j}{\sigma_{u_3}}\right)$$

$$= \beta_j - \frac{\psi_j}{\sigma_{u_3}}\beta_\lambda\lambda_i\left(\alpha_{u_3}\right)\left[\lambda_i\left(\alpha_{u_3}\right) - \left(\alpha_{u_3}\right)\right]$$

$$= \beta_j - \frac{\psi_j}{\sigma_{u_3}}\beta_\lambda\delta_i \qquad (2.9)$$

where $\delta_i = \lambda_i\left(\alpha_{u_3}\right)\left[\lambda_i\left(\alpha_{u_3}\right) - \left(\alpha_{u_3}\right)\right]$ This term is computed from the inverse Mills ratio λ_i to get $\delta_i = \lambda_i^2 - \lambda_i\left(X_{si}'\psi\right)$. Note $\phi\left(\alpha_{u_3}\right) = \lambda_i\left(\alpha_{u_3}\right)\left[1-\Phi\left(\alpha_{u_3}\right)\right]$ was used to get third step in the above equation (2.9). Since $0 < \delta_i < 1$, the additional terms serve to reduce the marginal effect.

In the case where an x_{ji} is a binary variable, the marginal effect will be expressed differently from the above result in equation (2.9). Let $\overline{X_{si}^{0}}$ be a vector of the mean values of the explanatory variables in the selection equation with $x_{ji}=0$ and $\overline{X_{si}^{1}}$ be the same vector of the explanatory variables in the selection equation but with the $x_{ji}=1$. Using equation (2.6), the conditional marginal effect of x_{ji} going from 0 to 1 is:

$$E\left(\Delta HIV_i \mid S_i^* > 0\right) = \beta_j + \beta_\lambda \Delta\lambda \tag{2.10}$$

where the marginal effect at the mean values are obtained by:

$$\Delta\lambda = \frac{\phi\left(\overline{X_{si}^{1'}}\psi/\sigma_{u_3}\right)}{\Phi\left(\overline{X_{si}^{1'}}\psi/\sigma_{u_3}\right)} - \frac{\phi\left(\overline{X_{si}^{0'}}\psi/\sigma_{u_3}\right)}{\Phi\left(\overline{X_{si}^{0'}}\psi/\sigma_{u_3}\right)} \tag{2.11}$$

2.5 Estimation Results

In this section we discuss the estimation results. Three models are estimated: the first model gives estimates without considering the sample selection and household wealth endogeneity bias. We used the set of dummy variables covering the quintiles for household wealth. The second model gives the estimates with correction for possible bias due the endogeneity of wealth, using IV probit estimation and the continuous wealth measure. In the third model, we use the Heckman procedure for the sample selection and the Rivers-Vuong procedure to deal with the endogeneity of the continuous household wealth variable. The following is a table giving the summary of equations used in the Models.

Table 2.1. **Summary of Equations in the Models**

Model	**Equation used (refer to section 2.3.4)**	**Explanation**
1	2.2	• Does not deal with endogeneity of wealth (exogeneity assumed) • Does not deal with HIV testing selectivity bias
2	2.2, 2.3 and 2.4	• Uses Instrumental variables to correct for wealth endogeneity • Does not deal with HIV testing selectivity bias
3	2.2, 2.3 and 2.5	• Uses wealth variable and its residuals predicted from regression of wealth on Instrumental variables as in 2.4 • A Heckman selectivity Model was fitted for child survival

2.5.1 Descriptive statistics

Table 2.2 gives a snapshot of HIV prevalence for a few socioeconomic sub-samples. Of those who agreed to be tested, women had a higher prevalence rate with 8.5% with HIV disease while 4.6% of the men had the disease. On average, out of the 6,360 individuals, only 422 (6.6%) individuals were found infected with HIV/AIDS. Women aged 15 – 19 years had higher rates of 8.9% and men with 4.6%. In terms of location, the disease is found to be more prevalent in the urban areas with 10.2% compared to the rural areas with only 5.6%. By province, Nyanza Province led with 14.3%, Nairobi taking the second place with 8.84% and the lowest being Eastern with 4.69%. With Kenya being composed of more than 70 ethnic groups, we found higher rates among Luo people with 22.6%, followed by Taita tribe with 9.6% and the least was Somali with 1.2%.

Using the data on household wealth given in quintiles the HIV prevalence rates was 3.1% for individuals who came from poorest quintile, 6.3% from the 20^{th} to 40^{th} percentile, 5.2% from the 40^{th} to 60^{th} percentile, 7.9% from 60^{th} to 80^{th} percentile and 10.3% from the wealthiest 20% of households.

In this study, we included a continuous household wealth in Models 2 and 3. To construct this continuous wealth variable we used a range of household assets to undertake principal components analysis, using the first component as the wealth index. As indicated earlier, the KMO value for the household assets was 0.914. KMO is a measure of the sampling adequacy used for comparison of the magnitude of the observed correlation coefficients and the magnitudes of the partial correlation coefficients. This value was well above the standard cut-off point of 0.6 which means the household asset variables used have much in common and hence it is worthwhile to use PCA. We used Bartlett's test of sphericity to test

the null hypothesis that the assets correlation matrix were uncorrelated. We were able to reject the null because of a high value for the computed chi-square, $\chi^2 = 1.21e+05$ and with 36 degree of freedom, the $p-value = 0.000$, indicating that the relationship among these household assets was very strong. The scree plot given in Figure 2.1 suggests we should pick one component since the elbow occurred at component 2. The eigenvalue for the first component was 4.725 and for component 2 was 0.98 which also indicates that it was enough to retain the first component.

2.5.2 Model 1: Probit Estimation Using HIV/AIDS Dependent Variable with Missing Data and No Endogeneity Correction

The estimates in this model used 5,490 individuals and the results are given in Table 2.4. The dependent variable is whether an individual is HIV positive given that he or she was from the selected household, was eligible for testing and agreed to testing. The estimation does not take into account for the possible endogeneity bias due to household wealth variable; in fact it uses the original indexed variable given in 5 wealth categories. The table gives the probit coefficients and marginal effects with robust standard errors.

Starting with the gender of household head, we find a negative significant coefficient of -0.176 and marginal effect of -0.018. This tells us that an individual will be 2% less likely to contract the HIV disease if the household head is male. The possible reason for this result could be that male-headed households are likely to have higher income and be more stable compared to female-headed ones. Individuals from female-headed households are more likely to be poorer, and more likely to engage into activities or jobs with high risks of obtaining HIV disease.

Looking at household wealth variables, they have mostly significant coefficients with marginal effects of 0.036, 0.035, 0.056 and 0.064 for poorer, middle, richer and richest household respectively. The marginal effects means that compared to the poorest household category, an individual from these households is more likely to contract HIV disease by between 3.6% and 6.4%. Indeed, as the household gets richer, the more likely the member from the household is to contract HIV. The possible reason behind these results could that individuals from wealthier household are able to afford to move places and socialise more, especially in the urban areas. In addition, males are able to pay for their sexual satisfaction more easily, which exposes them to greater risk.

An interesting result is found in the variable for gender, with a significant but negative coefficient of -0.258 and a marginal effect of -0.024 ($p<0.01$). This means a male is 2.4% less likely to acquire HIV compared to a female from the same household. This could be because of traditional cultures and customs where females are mostly under male authority and are less able to protect themselves due to their disadvantaged positions in sexual relationships. The polygamy variable has a significant coefficient of 0.221 ($p<0.05$) with marginal effect of 0.025. Considering the setup of Kenyan communities, their traditional cultures and customs make women more vulnerable to this kind of disease. Polygamy is practiced in Kenya, mostly in the rural areas, and we would expect it to contribute to the spread of HIV.

In the residential variables, we find a coefficient of 0.213 ($p<0.05$) with marginal effect of 0.023 if the household is located in urban areas. This suggests that an individual residing in an urban area is 2% more likely to get HIV disease compared to an individual residing in a rural area. Among the variables for provinces, there are only two provinces with significant

coefficients, Central and Nyanza provinces. These two have coefficient values of 0.302 ($p<0.05$) and 0.427 ($p<0.01$) with marginal effects of 0.035 and 0.053 respectively. There are several possible reasons as to why individuals in these provinces are more vulnerable to the disease. One of the reasons could be their high populations, which can be associated with many health issues. Due to high levels of wealth inequality, especially in developing countries, high population is often associated with economic difficulties and poor service delivery. In addition, for Nyanza province, the result could be partly explained by the lifestyle of the community occupying it. The community holds on to traditional cultures and customs making individuals, especially women, more vulnerable to HIV disease.

The results for ethnic groups show that only the coefficient for the Luo tribe is significant, with a coefficient of 0.719 and marginal effect of 0.110. This implies that an individual who is a Luo by tribe is 11% more likely to acquire HIV disease than an individual from the base tribe.[15] As mentioned in the case of Nyanza province, one possible explanation could be this community's traditional culture and customs. It could also be due to the Mombasa Kampala highway where we find heavy and long truck "trailers" transporting imported goods to the landlocked countries. The drivers and their co-drivers are thought to spread the HIV as they spend several days driving through.

The variable for an individual working permanently has an insignificant negative coefficient with a weakly significant marginal effect of -0.011. This means that an individual who is permanently working has about 1% lower chance of contracting HIV disease compared to an individual who is not working or works occasionally. Also included in the model is the individual's educational attainment. The variable for an individual who

[15] These includes other tribes not included in the model. Although there are about 70 ethnic groups, speaking different languages in Kenya, the model has included the main groups/tribes.

did not complete primary school has a coefficient of 0.226 with marginal effect of 0.023 ($p<0.1$). This implies that compared to an individual with no education, an individual with incomplete primary education is 2% more likely to acquire HIV. This is a surprising result. We expect little difference between "no education" (base) and incomplete primary education, although we note that the effect is barely significant.

The model also includes other individual variables including whether the individual's husband or partner lives away from the family. Surprisingly, this variable had a significant but negative coefficient of -0.233 and the marginal effect suggest that an individual whose husband or partner lives away is 2% less likely to acquire HIV. This is a surprising result because one would expect households where a partner is mobile to have a greater risk of contracting the disease. A cross tabulation of the variable for an individual's HIV status and the variable for the individual's husband or partner living away for the entire country and for urban and rural areas was done to confirm the above results. Although the Pearson chi-square statistics from the table indicate no significance, the probabilities for the entire country, urban and rural areas for husbands or partners living away and HIV positive are lower than for those living with their wives in the household (Table 2.3). This confirms the results in our models.

The variables for individuals' marital status show very strong effects. The marginal effects of the variables for marital status are all significant. These marginal effects tell us that an individual who is widowed has the highest risk of being infected with HIV, about 16% higher than the base of an unmarried person. Divorced individual follows with about 11%, then a separated individual with about 9%. An individual who is just living with the partner

has about 5% greater chance of being infected while a married individual has about 3% greater chance of acquiring the disease compared to an individual who is not married.

The variable for not using condoms during sex, although weakly significant, gave an interesting result with negative coefficient of -0.114 and marginal effect of -0.011. This means that compared to individuals using condoms, the individual who does not use a condom during sex is less likely to get infected with HIV. This is not what we expected since use of condoms would be protecting an individual from contracting the HIV disease or even any other sexually transmitted diseases. However, the effect is weak, so may not be a robust result. The variable for an individual having multiple sexual partners has a significant coefficient of 0.133 ($p<0.05$) and marginal effect of 0.013. This means that compared to an individual who has only one sexual partner, the individual with many partners is 1.3% more likely to contract HIV, as expected.

The last variable in the model is for male circumcision which has a coefficient value of -0.237 and marginal effect of -0.020 ($p<0.05$). This implies that male circumcision reduces the chances of an individual contracting HIV by about 2%.

2.5.3 Model 2: IVProbit Estimation Using HIV/AIDS Dependent Variable Endogeneity Correction

The results discussed above do not account for the fact that wealth is likely to be endogenous, and that the sample includes potential selection bias in who agrees to be tested. The model results shown in column 2 of Table 2.4. deals with the endogeneity of wealth, by use of Instrumental Variables. A set of instruments are included in the equation

for household wealth. First we discuss the results for the HIV/AIDS equation and then the results for the IV equation on wealth.

Compared to Model 1, Model 2 gives mostly similar results, with a few differences. Model 2 has a continuous household wealth variable which is found to be significant with a coefficient of 0.122 and marginal effect of 0.012 ($p<0.1$). This means that for one a unit increase in the wealth index, an individual will be 1.2 percentage points more likely to be infected by HIV disease. This is consistent with the wealth quintile results from Model 1.

The main difference observed in this model is in the variables for provinces where all provinces apart from the Coast province have significant coefficients. Still the Nyanza province is strongest where an individual living in the province is about 10% more likely to acquire the HIV disease compared to an individual living in Nairobi province, a much stronger effect than when the endogeneity of wealth is not dealt with.

There are other small differences compared to Model 1. First, an individual who is aged 15 to 35 years is now weakly significant with marginal effect of 0.014 meaning an individual is 1.4 percentage points more likely to have HIV if they are aged between 15 and 35 years compared to those aged between 36 and 49 years. Secondly, the male circumcision variable is now insignificant.

Table 2.5 gives the results for the endogenous wealth variable equation (IV Model). There are two excluded instruments with significant coefficients. Both variables capture the effect of individual's occupation: the variable for self employed in agriculture had negative coefficient of -0.455 ($p<0.1$). This means that an individual who is self-employed in the agriculture work has lower wealth. The other variable is for those with a household and

domestic occupation, with a coefficient of 1.737 ($p<0.01$). This suggests that an individual working in the household and domestic is more likely to be rich as compared to an individual whose occupation is unknown.

A number of variables affect individual's wealth. An individual from a household headed by a male is more likely to be wealthier as in the case if that person is a male. An individual working permanently has higher wealth as is the case if that person had some education. A person aged between 15 and 35 years has lower wealth as does the person who is married, just living together, widowed, divorced or separated. Those who do not use condoms have lower wealth as do those who have many sex partners. It is surprising to find that an individual who is a prostitute has more wealth. Also, male circumcision is associated with lower wealth.

The Wald test of exogeneity of the instrumental variables is insignificant indicating no endogeneity or that the instruments used were weak. It is important to note that an instrument is relevant if its effect is statistically significant, is strong if the size of its effect is large and valid if it is uncorrelated with the structural error term (Mwabu, 2009). However, it is not easy to get an instrument satisfying these conditions (Bound *et al.*, 1995). According to Mwabu (2009), if the instruments are relevant but weak, then use of 2SLS estimator is biased towards the OLS estimator giving potentially biased estimates.

In this study, using two step estimation, the joint *p*-value and F statistics on the excluded instruments showed that the instruments for this model are relevant (result not given). The F statistic gives vital information as to the validity and relevance of instruments in the case of a single endogenous variable (Shea, 1997). Our model has a single endogenous variable and the F statistic on excluded instruments can be used to determine whether or not the

relative bias of the IV estimates in the presence of weak instruments is sufficiently small (Stock *et al.*,2002). In this study, the F-statistics on instruments for the input equations are low, suggesting that the instruments may be weak.

2.5.4 Model 3: Heckprob Estimation: Bias correction for Sample Selection and household wealth variables

In Model 3, we correct for both the possible bias caused by the non-random selection of individuals who agreed to be tested for HIV and the endogeneity of household wealth. To correct for sample selection bias the Heckprob procedure was used. The selection variable was whether the individual agreed to be tested or otherwise, having been selected for HIV testing.

Model 3 uses 6,941 individuals with 1,752 censored individuals and the estimates are given in Table 2.4 *column 3*. The table shows the probit coefficients and their respective marginal effects with robust standard errors.

The results in this model are very similar to those in Model 1 and 2. In some cases variables go from being marginally insignificant to marginally significant (eg. The variable for if a person uses a condom during sex). The other noticeable difference is on the magnitudes and significance of the marginal effects. Most of the marginal effects have decreased slightly, meaning weaker effects. Examples of such results include the provincial variables where the marginal effects of all provinces have decreased, except the province of Nyanza. The estimates are similar in magnitude to those in the other two models. Overall, though the marginal effects have decreased considerably compared to those in Model 2, reflecting the loss of information when selection is accounted for.

2.5.4.1 Heckman Correction (Sample Selectivity) Model

The estimates of the selection equation are given in Table 2.5 column 2. In this equation, three instrumental variables are included, along with exogenous independent variables from the main model. The results show that the variables for an individual whose interview was interrupted is significant with a coefficient of -0.176 ($p<0.1$). This implies that compared to an individual who had no interruption during the interview, an individual who had the interview interrupted by any of the adult household members was less likely to refuse or ignore the HIV testing. Another significant instrument is for an individual being a Protestant or Seventh Day Adventist. The variable had a coefficient value of 0.254 ($p<0.1$) implying an individual who was a Protestant or Seventh Day Adventist was more likely to agree to be tested for the disease compared to an individual who belonged to other religion or no religion.

There are other variables that influence the individual's choice to be tested. Results suggest that individuals from the urban areas were less likely to be tested compared to individuals from the rural areas. In contrast, an individual from the Nyanza and Western province were more likely to be tested for the disease compared to an individual from the Nairobi province. In addition, an individual who was a Kalenjin or Kisii by tribe was more likely to be tested. However, the coefficient for a Mijikenda/Swahili individual showed that such an individual is less likely to be tested.

The variables for a widowed individual and a prostitute showed they are more likely to be tested. But the coefficient for an individual who did not abstain from sex showed this individual being less likely to be tested compared to the one who abstains from sex. Also,

male circumcision is associated with an individual being less likely to be tested for the HIV disease.

The Likelihood Ratio test is used to test for the errors of the two equations being independent (main and selectivity). The *p*-value ($p = 0.091$) indicates some evidence for the presence of unobservables that are correlated between the selection and the main equation and so there would be bias in the estimates if we ignored the sample selection issue.

2.6 Chapter Discussion and Conclusions

In the preceding section we present models estimating the factors which might contribute to individuals being infected by HIV/AIDS. Although the discussion in this section is based on the results from the final model that accounts for endogeneity of wealth and sample selection, we do refer to Model 1 in a few cases. Most previous studies have not considered the possible bias due to endogeneity of household wealth or income.

This study finds that gender of the household head contributes to an individual's HIV status, with individuals from female-headed households being more likely to be HIV positive. The relative poverty of female-headed households is usually linked to women being disadvantaged with respect to either assets or/and activities, linked to inequalities of access to resources as well as income generating opportunities, especially with high rates of unemployment in developing countries (Ellis, 2000. see also Chant, 2003; Staten, *et al*., 1999). In some cases, the female headship arises because of death of the spouse where the female is forced to head the family. There are many factors associated with disadvantage of females who in turn become heads of their households. For example, Fukuda-Parr (1999)

indicated that unequal opportunity in schooling for girls and boys restricts choices not only in employment but also for a creative life. Most of literature on household headship in Africa has focused on the well-being of the household and the children where they show the weakness and vulnerability of female household headships (for example Desai, 1992; Posel, 2001).

The results show that household wealth plays a major role in spreading HIV disease. As the household becomes wealthier, individuals from the household are more vulnerable to catching the disease. This is not necessarily what might have been expected. The possible reason explaining these results could be that individuals from wealthier households are more mobile and as a result more likely to have multiple partners and therefore more likely to engage in activities that expose them to higher lifetime HIV risks.

There are mixed findings in earlier literature on how household wealth affects the prevalence of HIV. The study by Shelton, Cassell and Adetunji (2005) showed that poverty has a varying effect on the HIV epidemic. Using data from Tanzania, they found household wealth was strong positively related to HIV prevalence. There was a four-fold difference in prevalence for women between the lowest and highest wealth quintile, with the highest wealth quintile more likely to be HIV positive (see also Tanzania Commission for AIDS [TACAIDS], 2005). In fact, in a recent study, Parkhurst (2010) with evidence from Tanzania, indicated that both wealth and poverty can lead to potentially risky or protective behaviours which are responsible for HIV growth. Parkhurst found that the relationship between wealth and HIV infection can change over time in a given setting, with declining prevalence in wealthy groups occurring simultaneously with increasing prevalence in poorer women.

In a study based in Blantyre in Malawi, women with husbands of high socio-economic status were more likely to be HIV infected compared with women with lower socio-economic status husbands (Taha, *et al.*, 1998). The study done by Gillespie, Kadiyala and Greener (2007) indicated that most studies focus on relative poverty in the context of generalized chronic poverty. In their study, although there was a weak positive relationship between national wealth and HIV prevalence across countries in sub-Saharan Africa, there was a clear and significant pattern of association between income inequality and HIV prevalence across countries. The study by Mishra, *et al*. (2007; 2009) indicated HIV prevalence to be higher among people with more household wealth. In a study in Africa from 2002–2005, Dinkelman, Lam and Leibbrandt (2007) indicated that for girls, sexual debut appears to be earlier in poor households, especially those who have experienced an economic shock because of a death, illness or job loss of their household earner. Another study by Rowley, *et al*. (2008) showed that limited access to income would be a factor to put individuals into high risk of becoming HIV positive.

Our result confirms the earlier findings that household wealth contributes directly and indirectly to the spread of HIV disease among individuals. As most of the factors have been included in our models, other individual and household factors linked to wealth act as a catalyst towards exposing individuals to high risks of acquiring the disease.

Our study shows gender differences in the risk of acquiring HIV, with males being less likely to be infected. This could be due to many factors such as the relative ease of income and opportunities for males, as well traditional cultures and customs around the treatment of women. Earlier studies on this topic in Kenya records similar results. Using a small sample among agricultural plantation residents in Kericho, Sateren *et al.* (2006) got an

overall HIV-1 prevalence of 17.4% for women, more than twice that of men (8.0%). Similar results were found by Rowley, *et al*. (2008) where across different age groups they found the gender difference for all the age groups to be significant with the biggest difference for the 15 – 24 year age group.

The results from this study established that the practice of polygamy increases the chances of an individual acquiring the HIV disease compared to a monogamous family. This is expected when one looks into the issue of traditional cultures and customs where in most cases there is discrimination against women, sometimes leaving them powerless under a male's authority. As a cultural practice, polygamy is common in most communities in Kenya even today. The practice has declined in recent years due to the influence of foreign or western culture, Christianity, education and pressure from economic constraints mostly in the urban areas. In the rural areas, many communities still uphold the practice, and it is clearly a pathway to spread the HIV disease. Women are unable to refuse to have sex with their husbands.

In their study on HIV/AIDS in sub-Saharan Africa, Lau and Muula (2004) indicated that many women are forced into a new marriage, probably after their partner dies, a practice adding to health risks. One example of this kind of lifestyle is found among the Luo tribe where a wife is inherited by one of the brothers or a close relative if her husband dies and she is expected to meet all her marital requirements. Inherited women may become infected with HIV and eventually die of AIDS, leaving children orphaned. In other cases, after death of the husband, the woman is believed to have acquired contagious cultural impurity that is considered dangerous to other people. The community believe that this impurity can be cleared by performing a sexual cleansing ritual where the widow has sex with a brother-in-

law or cousin of the deceased husband, and in case they cannot, with a professional cleansing man. This kind of act could be a tool for HIV/AIDS transmission to the widow or other people in the community (Ambasa-Shisanya, 2007; Centre for Rights, Education and Awareness [CREAW], 2008; Kenya Information Guide, 2010; Amnesty International Document, 2002). This is mostly found in the rural community where these traditional cultures and customs hold fast. [16]

Our study confirms that polygamy together with other traditional and cultural practices are more likely to expose individuals to health risks such as acquiring HIV and other diseases. Although these practices are part of the community norms, the risks involved need to be highlighted to communities. On the other hand, although traditional practices do contribute to high prevalence of HIV, this is mainly observed in specific areas especially the Nyanza province.

From the descriptive statistics, HIV prevalence was found to be higher in the urban than in the rural areas. The analysis confirms that an individual living in the urban areas is at higher risk of contracting HIV than in the rural areas. There are various factors contributing to this effect including wealth or income inequality, high population, health problems, diversity of cultures, etc.

At the provincial level, an individual from the Nyanza and Central provinces had highest risk of getting HIV disease. This high prevalence of HIV/AIDS in Nyanza is noted in earlier studies (see National AIDS/STD Control Programme [NASCOP], 2005; Lau & Muula, 2004). Among the reasons for such results could be high population density in these provinces. As in the urban areas, these two provinces are highly populated. The Central

[16] For more see: WUNRN, 2006; BBC News, 2003

province surrounds Nairobi province (literally the Nairobi city) and is composed of most ethnic tribes in the industrialised towns like Thika and Nyeri. High populations are basically associated with low income and lack of health facilities making the spread of HIV disease more vigorous. The United Nations report (United Nations [UN], 1996) indicates that in some cities, including Nairobi, more than half of the population live in slums and squatter settlements. Due to continued unemployment and underemployment the people living in the slums live in poverty, with insufficient incomes to fulfil even basic nutritional and shelter requirements. In urban areas, there is close proximity and frequency of interaction among diverse groups of people that speeds up the rates of HIV infection and transmission and the treatment for infected individuals is a problem. In addition, diagnosing the disease in such areas is a challenge and in most cases is undiagnosed until an individual becomes very sick or the symptoms become more visible, which puts the uninfected in danger. The United Nations Human Settlement Programme [UN–HABITAT] (2002) report highlights such problems associated with densely populated areas.

Apart from the high populations, Nyanza is occupied by the community called Luo. This community practice the traditional cultures and customs that put individuals at high risk of infection and spreading of HIV disease. One example of such practices is wife inheritance, and their belief that when a man dies, the wife has to undergo a cleansing ceremony called "tero buru" to cast the evil spirits away. The ceremony involves a sexual component, before the woman can be re-incorporated into society (Luginah, *et al.*, 2005; Ambasa-Shisanya, 2007; Ayikukwei, *et al.*, 2007). This ritual endangers widows' lives through possible infection with HIV/AIDS.

In addition, male individuals from the Luo community are usually not circumcised. This worsens the situation because if the uncircumcised cleansers are infected, there is a high chance of transferring the HIV disease to the widows. This is supported by the results from our study in Model 1 where male circumcision is found to reduce chances of an individual getting HIV disease. Previous literature indicates that male circumcision protects against HIV infection. Also Sateren *et al.* (2006) found that among the Luo tribe respondents there were about 23.5% with HIV, with that rate for uncircumcised being about 29.2%.[17] The reason for this could be that uncircumcised males suffer bruises during intercourse and this exposes their partners to infection more readily than the circumcised (Scott, Weiss & Viljoen, 2005). In their study, Szabo and Short (2000) give the medical and physical evidence on male circumcision and HIV infection.

In addition to this, another factor responsible for spreading HIV is the heavy traffic of the long distance trucks from Mombasa port to landlocked countries including Uganda, Burundi, Rwanda, Democratic Republic of Congo and other parts of Kenya via the Mombasa – Kampala Highway. On the Highway, there are stops on the route where the drivers stop their trucks overnight. The truck drivers and their co-drivers or conductor remain highly mobile and spend a long time, even several weeks, in transit away from their families. Their need for female companionship makes them more likely to use the services of commercial sex workers in stopover towns. Networks and services meeting the business and recreation needs of these truck drivers have developed, including gas stations, inspection points, lodges, bars and brothels, and they have a high population of commercial sex workers. Often, the commercial sex workers are driven to this type of work due to financial difficulties and poverty in their households and this being the easy way to make

[17] See also Moses *et al.* (1990)

money, they become highly vulnerable to HIV disease. As a result the whole lifestyle becomes a key factor to spread the disease to even the rural communities when the victims travel up country to visit their family members. At the border of Kenya and Uganda, due to clearance of the duty and customs of the goods leaving Kenya, these trucks may spend several days in the queue before crossing the border. The Luo community occupies this border region.

Several empirical studies have established a link between human mobility and the risk of HIV transmission. In sub- Saharan Africa, the risk of HIV infection has been found to be higher near roads, and among individuals who either have personal migration experience or have sexual partners who are migrants (for example Hudson, 1996; Bloom, *et al.*, 2002; Zuma *et al.*, 2003; Boerma *et al.*, 2003; Lagarde *et al.*, 2003; Barnighausen, *et al*., 2007). In Eastern Africa, several studies reported HIV prevalence rates of 25% to 32% among truck drivers in Kenya and Uganda (Mbugua *et al.*, 1995; Bwayo *et al*., 1994, see also Morris & Ferguson, 2006). A study in 1997 of 200 adolescents aged 15–19 years of age at truck stops in Kenya showed that 46% engaged in sex with transport workers, 78% of the females had traded sex for gifts or money, and 52% had experienced a sexually transmitted disease (Nzyuko *at el*., 1997).

Another crucial factor which could explain results for provincial and ethnic tribes contributing to individuals acquiring HIV is the lack of knowledge and education on HIV in the communities, especially in the rural areas. Inadequacy of other interventions such as free distribution of condoms, AIDS education in schools, voluntary HIV testing and counselling could be contributors to the individuals in these provinces acquiring HIV disease. The provision of such facilities could reduce the incidence of high rates of the

disease in the country. A follow up study done by Lugalla *et al.* (2004) on social, cultural and sexual behavioural factors, indicated a decline in HIV infections trends in Kangera region in Tanzania. Due to interventions such as health education, distribution of condoms, education of HIV in schools and provision of counselling and testing centers, the infection of HIV trends was noticed to have declined. In addition, the paper notes that the decline in HIV trends in the area was a partly due to decline of traditional practices such as polygamy, widow inheritance, excessive alcohol consumption, and sexual networking.[18]

The results in our study found that working permanently reduces the chances of an individual acquire the disease. This could be because the majority of workers belong in middle class occupations with moderate income. These individuals have enough to support their families and struggle to meet their needs including health, education, clothing, housing etc. In their study Bowen *et al.* (2008) found that the lowest class of people are often associated with the lowest occupations, part-time or casual jobs where the majority are not trained. Casual labourers had the highest prevalence rate of HIV/AIDS followed by temporary employed general labours while permanently employed, semi-skilled operators and drivers and skilled employees were next.

Our study showed that marital characteristics play a big role in putting the individuals into risk of getting HIV disease. A widowed individual was found to have the highest risk followed by divorced, then separated and the lowest risk is for a married individual. The possible explanation for these results could be that a widow or divorced or separated female does not get any income support from the husband or male partner anymore. In addition, such individual is psychologically affected by the loss of the partner. These and other

[18] See also Sheton & Johnston, (2001).

factors may force an individual to do anything to acquire income to support her children or other family members. Big numbers of them will engage into jobs which may risk them acquiring the HIV disease. Furthermore, if the deceased husband or partner died due to HIV/AIDS, then the wife would probably be infected too. The study done by Carson *et al.* (1998) gave similar results. They found that women had a higher infection rate than men and those who worked as bargirls or were divorced, widowed or separated were particularly at risk. Another study by Watson-Jones *et al*. (2009) found that HIV incidence (per 100 person-years) was 3.6 for unmarried, 3.88 for married and 4.81 for divorced, separated or widowed. Also a study by Mishra, *et al*. (2009) showed individuals who were widowed and divorced had greatest risk while those who had never married were at lowest risk.

There is indirect evidence from Blantyre in Malawi, to support the notion that women with husbands of high socio-economic status were more likely to be HIV infected compared with women with lower socio-economic status spouses (Lau & Muula, 2004). In developing countries, such as Africa, the traditional gender roles leaves the female with little control over her sex life and men are taken to be very knowledgeable. Migratory practices in search of jobs in the urban areas force spouses to be separated for extended periods of time causing extramarital relationships. Due to this situation the risk of HIV is not considered an important issue compared to day-to-day survival (Lau & Muula, 2004; Buve, *et al.* 2002). In his study on the determinants of HIV infection and sexual behaviours in Burkina Faso, Cameroon, Ghana, Kenya and Tanzania, De Walgue (2006) found that married women who engage in extra-marital sex are less likely to use condoms than single women when doing so. In addition, he found having been in successive marriages is a significant risk factor, as evidenced by the results on HIV infection and on sexual behaviours.

The results show that having multiple sex partners poses a high risk of an individual contracting HIV. In Africa, the spread of HIV disease is mostly through sexual relationships. Most of the health messages of abstinence and being faithful to your partners are often cited as methods of HIV prevention. The study by Nzyuko, Morgan and Nyamwaya (1991) showed that about 73% of the population surveyed said HIV could be prevented by sex with one partner, abstinence or no extra-marital sex. In the study by Bwayo *et al.* (1991), about 67% of truck drivers said faithfulness to your partner and avoiding sex with prostitutes could prevent the spread of HIV. A recent cross-sectional study in Kenya found asset poverty to be significantly related to risky sexual outcomes, such as early sexual debut and multiple sexual partnerships, in all three residential settings studied (Nii-Amoo Dodoo, Zulu & Ezeh, 2007).

Our study shows that the use of condoms increased the chance of contracting HIV. This is not what we expect. It is commonly believed that use of condoms reduces the spread of the disease (Lugalla *et al.,* 2004). In the earlier studies using Kenyan national Demographic and Health Surveys for 1993, 1998, and 2003, it was found that use of condoms was rising, leading to a decline of HIV prevalence (Cheluget *et al*., 2006; UNAIDS, 2006). However, their study does not investigate whether the use of condom reduces the incidence of HIV, but it does confirm that there was an increase of condom use. In other studies on Kenyan truck drivers, only 17% (Bwayo *et al.*, 1991) and 11% (Nzyko, *et al.*, 1991) cited condoms as a method of preventing HIV transmission. However, the reason for the results in our study is not well understood. The reason behind our results could be due to an issue of consistency in use of condoms. There are a number of studies which have found inconsistent users of condoms are at higher risk than never users. In their study, Ahmed, *et al.,* (2001) found only 4.4% reported consistent condom use and 16.5% reported

inconsistent use during the prior year. This irregular condom use was found not to be protective against HIV or STD and was associated with increased gonorrhea/Chlamydia risk. Similar results were found by Deschamps, *et al.*, (1996) where they found the incidence in sexually active couples who infrequently used or did not use condoms was 6.8 per 100 person-years. Taha *et al.*, (1996) reported that although consistent condom use peaked at 62% in the first 6 months, it declined to as low as 8% in the second year of follow-up. Condom use at each visit, either intermittent or consistent, was higher in HIV-seropositive than HIV-seronegative women. It might also be that some people who report always using condoms do not actually do so but tell the researchers that they do in order to "look good". Similar points could be made about people reporting not using condoms. Self-reported data always have the risk of being unreliable; and in this case there is no other means of confirming the findings. Somewhere it was noted in the speech by one of the Kenyan leaders saying that use of the gadget "*condom*" was causing the spread of AIDS in the country (Smith, 2006). This is consistent with the findings in this study.

Potentially the findings in this chapter are of value to both researchers and policy makers.

The factors found in this study focus on culturally-defined gender values and norms that evolve through a process of socialization starting from an early stage of infancy. They determine and reinforce themselves through traditional practices such as wife sharing, widowhood related rituals, early marriage, female genital mutilation and the condoning of gender-based violence which strongly influence the visible aspects of individual behaviours and are important determinants of individual's vulnerability to HIV, especially women.

Previously there have been strategies campaigning on control and prevention measures in reduction of HIV/AIDS prevalence to individuals of all age groups. Improved targeting of

such policies or programmes would further reduce the number of people contracting HIV/AIDS. An example of such is to educate women because the study found they are more vulnerable due to tradition and cultural norms. In addition, the focus should be directed to the highly affected regions including rural areas such as Nyanza province, slums in the urban areas, etc, where the study finds high prevalence of HIV/AIDS. Generally, the study found the urban areas had highest prevalence rates. One of the urban areas to focus on is the slums, as it is recognised that the HIV disease is more prominent in the slums. Use of new, evidence-based approaches to HIV prevention will to a great extent help to reduce the spread of HIV. These may include the availability and access to counselling and testing and distribution of condoms free of charge to the commercial sex workers and with subsidised cost at the shops and hotels. Ensuring more effective and targeted behaviour change communication, promoting abstinence, safe sex and delayed sex debut among young people would bring huge reduction of HIV prevalence. Expanding services for prevention of mother-to-child transmission and improving availability of safe blood supplies, ensuring injection safety and expand access to post-exposure prophylaxis and universal precautions will as well be important intervention in reduction of spreading the disease. The findings in this study are potentially very useful in contributing to the research field as well as to the policy makers in the enhancement of their strategies for these HIV/AIDS interventions.

Table 2.2. Descriptive Statistics for Variables contributing to HIV/AIDS

Variable	Sample size	Proportions (Std Dev)[19]				
		By type of place of residence			*By gender*	
		Overall	Urban	Rural	Female	Male
Individual with HIV/AIDS	6360	0.066 (0.249)	0.102 (0.302)	0.056 (0.229)	0.0846 (0.278)	0.0463 (0.210)
		By Household Wealth Quintile				
		Poorest	Poorer	Middle	Richer	Richest
		0.031 (0.172)	0.063 (0.243)	0.052 (0.222)	0.079 (0.269)	0.103 (0.304)

Variable	**Sample size**	**Other Variables Proportions**	**HIV proportions by other variables**
Provinces			
Nairobi	31736	0.108 (0.311)	0.089 (0.284)
Central	31736	0.150 (0.356)	0.054 (0.227)
Coast	31736	0.116 (0.320)	0.065 (0.248)
Eastern	31736	0.126 (0.332)	0.049 (0.217)
Nyanza	31736	0.123 (0.329)	0.143 (0.351)
Rift Valley	31736	0.181 (0.386)	0.048 (0.215)
Western	31736	0.121 (0.327)	0.056 (0.230)
North Eastern	31736	0.073 (0.261)	0.0
Ethnic group			
Embu	31736	0.011 (0.102)	0.020 (0.142)
Kalenjin	31736	0.089 (0.285)	0.036 (0.186)
Kamba	31736	0.097 (0.296)	0.067 (0.251)
Kikuyu	31736	0.211 (0.408)	0.051 (0.220)
Kisii	31736	0.051 (0.221)	0.044 (0.205)
Luhya	31736	0.145 (0.352)	0.063 (0.243)
Luo	31736	0.101 (0.301)	0.226 (0.418)

[19] Standard Deviations in parenthesis

Table 2.2. ***Continued***

Variables	Sample size	Proportions (Std Dev) Other Variables Proportions	HIV proportions by other variables
Masai	31736	0.026	0.036
		(0.160)	(0.186)
Meru	31736	0.042	0.047
		(0.202)	(0.212)
Mijikenda/Swahili	31736	0.075	0.049
		(0.264)	(0.216)
Somali[20]	31736	0.094	0.012
		(0.292)	(0.108)
Taita	31736	0.015	0.096
		(0.122)	(0.297)
Turkana	31736	0.020	0.041
		(0.139)	(0.199)
Kuria	31736	0.007	0.022
		(0.082)	(0.147)
Other tribes	31736	0.015	0.052
		(0.122)	(0.223)
Religion			
Catholic	31703	0.233	0.067
		(0.423)	(0.250)
Protestestant or Seventh Adventist	31703	0.614	0.075
		(0.487)	(0.264)
Muslim	31703	0.130	0.034
		(0.336)	(0.181)
No religion	31703	0.019	0.066
		(0.137)	(0.250)
Other religion	31703	0.004	0.000
		(0.064)	(0.000)
Occupation			
Not working or not known	31689	0.435	0.063
		(0.496)	(0.243)
Professional or Technical or Managerial	31689	0.030	0.072
		(0.170)	(0.259)
Clerical	31689	0.010	0.089
		(0.100)	(0.288)
Sales	31689	0.136	0.081
		(0.343)	(0.274)
Agriculture – Self Employed	31689	0.280	0.056
		(0.449)	(0.231)
Household and Domestic	31689	0.054	0.105
		(0.227)	(0.307)
Services	31689	0.016	0.141
		(0.124)	(0.350)
Skilled Manual (skilled & unskilled)	31689	0.054	0.099
		(0.226)	(0.299)
Education			
No education or Pre-Primary school	31736	0.184	0.035
		(0.388)	(0.184)
Incomplete Primary education	31736	0.326	0.076
		(0.469)	(0.265)

[20] Somali occupy North eastern province which had a small sample and no HIV victims. This percentage for Somali indicates the Somali individuals who have HIV/AIDS but do not live in the province. For example, there is a big population of Somali community living in Nairobi city.

Table 2.2. ***Continued***

Variables	Sample size	Proportions (Std Dev) Other Variables Proportions	HIV proportions by other variables
Primary level	31736	0.233	0.081
		(0.423)	(0.273)
At least Secondary education	31736	0.211	0.068
		(0.408)	(0.252)
Higher education	31736	0.046	0.059
		(0.209)	(0.236)
Marital Status			
Never married	31736	0.301	0.046
		(0.459)	(0.210)
Married	31736	0.560	0.063
		(0.496)	(0.244)
Just living with partner	31736	0.051	0.100
		(0.220)	(0.301)
Widowed	31736	0.034	0.193
		(0.181)	(0.396)
Divorced	31736	0.016	0.138
		(0.126)	(0.348)
Separated	31736	0.038	0.154
		(0.190)	(0.362)
Others			
First sex intercourse below 15 years of age	31683	0.228	0.107
		(0.420)	(0.310)
No sex abstain	31729	0.494	0.071
		(0.500)	(0.258)
No condom use during sex	31729	0.518	0.052
		(0.500)	(0.222)
Many sex partners	31729	0.364	0.073
		(0.481)	(0.260)
Prostitution	31729	0.846	0.071
		(0.361)	(0.257)
No injection avoidance	31729	0.834	0.072
		(0.371)	(0.258)
Household member gender	37611	0.496	0.046
		(0.500)	(0.210)
Interview interruption	37612	0.38	0.086
		(0.191)	(0.280)

Table 2.3. **Cross tabulation for HIV individuals by Husband live away**

		Husband live away		
		No	Yes	Total
HIV status	Negative	4,612 (93.08%)	515 (93.98%)	5,127 (93.17%)
	Positive	343 (6.92%)	33 (6.02%)	376 (6.83%)
	Total	4,955 (100.00%)	548 (100.00%)	5,503 (100%)
		Pearson $\chi_1^2 = 0.628$, $p = 0.428$		
		In the Urban areas		
	Negative	3,336 (94.37%)	421 (95.03%)	3,757 (94.44%)
	Positive	199 (5.63%)	22 (4.97%)	221 (5.56%)
	Total	3,535 (100.00%)	443 (100.00%)	3,978 (100.00%)
		Pearson $\chi_1^2 = 0.330$, $p = 0.566$		
		In the Rural areas		
	Negative	1,276 (89.86%)	94 (89.52%)	3,757 (94.44%)
	Positive	144 (10.14%)	11 (10.48%)	155 (10.16%)
	Total	1,420 (100.00%)	105 (100.00%)	1,525 (100.00%)
		Pearson $\chi_1^2 = 0.012$, $p = 0.913$		

Table 2.4. **Models 1 - 3: Probit, IVprobit and Heckman Estimates for HIV/AIDS Individuals**[21]

Dependent Variable: Individual with HIV/AIDS						
	Probit Model (1)		**IVProbit Model (2)**		**Heckprobit Model (3)**	
	Number of observations = 5,490		Number of observations = 5,195		Number of observations = 6,941	
	Wald χ^2 = 348.17		Wald χ^2 = 318.85		Censored observations = 1,752	
	Prob > χ^2 = 0.000		Prob > χ^2 = 0.000		Uncensored observations = 5,189	
	Degrees of freedom = 41		Degrees of freedom = 38		Wald χ^2 = 270.84	
					Prob > χ^2 = 0.000	
					Degrees of freedom = 39	
Variables	**Coefficient**	**Marginal Effect**	**Coefficient**	**Marginal Effect**	**Coefficient**	**Marginal Effect**
Household Characteristics						
Gender of household head (*Base-Female*)	-0.176**	-0.018**	-0.190**	-0.020*	-0.142*	-0.016
	(0.082)	(0.009)	(0.093)	(0.011)	(0.084)	(0.010)
Poorer household (*Base – Poorest houseld*	0.311**	0.036**				
	(0.128)	(0.017)				
Middle household	0.304**	0.035**				
	(0.126)	(0.016)				
Richer household	0.459***	0.056***				
	(0.126)	(0.019)				
Richest household	0.525***	0.064***				
	(0.154)	(0.023)				
Wealth, Continuous			0.122*	0.012*	0.124**	0.013**
			(0.067)	(0.007)	(0.056)	(0.006)
Wealth Continuous residuals					-0.092	-0.010
					(0.062)	(0.006)
Household member gender (*Base-Female*)	-0.258***	-0.024***	-0.301***	-0.029***	-0.262***	-0.028***
	(0.064)	(0.006)	(0.065)	(0.006)	(0.069)	(0.006)
Polygamy	0.221**	0.025*	0.201*	0.023	0.207*	0.025*
	(0.111)	(0.014)	(0.112)	(0.015)	(0.108)	(0.015)

[21] Probit gives estimates without endogenous bias correction which is corrected in IVprobit. Indexed (ordinal) household wealth variable is used and HIV sample section not taken into account.

Table 2.4. ***continued***

Dependent Variable: Individual with HIV/AIDS

	Probit Model (1)		IVProbit Model (2)		Heckprobit Model (3)	
Variables	**Coefficient**	**Marginal Effect**	**Coefficient**	**Marginal Effect**	**Coefficient**	**Marginal Effect**
Residence						
Urban (*Base- rural*)	0.213**	0.023*	0.220*	0.023*	0.293**	0.019
	(0.104)	(0.012)	(0.120)	(0.014)	(0.134)	(0.016)
Coast province (*Base - Nairobi*)	0.108	0.011	0.193	0.022	-0.031	0.019
	(0.175)	(0.020)	(0.183)	(0.024)	(0.219)	(0.022)
Central province	0.302**	0.035*	0.459***	0.060**	0.408**	0.055**
	(0.150)	(0.021)	(0.168)	(0.029)	(0.162)	(0.026)
Eastern province	0.211	0.023	0.465**	0.062*	0.338*	0.055**
	(0.159)	(0.020)	(0.184)	(0.032)	(0.185)	(0.028)
Nyanza province	0.427***	0.053**	0.666***	0.098**	0.421*	0.100***
	(0.144)	(0.023)	(0.191)	(0.040)	(0.247)	(0.036)
Rift valley province	0.204	0.022	0.401**	0.050*	0.269	0.044*
	(0.146)	(0.018)	(0.181)	(0.028)	(0.176)	(0.024)
Western province	0.208	0.023	0.453**	0.059	0.241	0.054*
	(0.154)	(0.019)	(0.219)	(0.038)	(0.232)	(0.032)
Individual Characteristics						
Tribe						
Kalenjin (*Base - Other*)	0.047	0.005	0.082	0.009	-0.023	0.009
	(0.194)	(0.020)	(0.198)	(0.022)	(0.197)	(0.022)
Kamba	0.210	0.024	0.201	0.023	0.177	0.023
	(0.141)	(0.018)	(0.147)	(0.019)	(0.138)	(0.019)
Kikuyu	-0.002	-0.000	0.108	0.011	0.124	0.012
	(0.146)	(0.014)	(0.149)	(0.016)	(0.136)	(0.016)
Kisii	-0.066	-0.006	0.015	0.001	-0.154	-0.002
	(0.191)	(0.017)	(0.199)	(0.020)	(0.212)	(0.019)
Luhya	0.116	0.012	0.195	0.022	0.093	0.019
	(0.159)	(0.017)	(0.168)	(0.021)	(0.173)	(0.020)
Luo	0.719***	0.110***	0.752***	0.120***	0.618***	0.112***
	(0.147)	(0.032)	(0.155)	(0.034)	(0.178)	(0.033)
Mijikenda/Swahili	0.095	0.010	0.192	0.022	0.292	0.014
	(0.232)	(0.026)	(0.243)	(0.032)	(0.230)	(0.029)

Table 2.4. ***continued***

Dependent Variable: Individual with HIV/AIDS

	Probit Model (1)		IVProbit Model (2)		Heckprobit Model (3)	
Variables	**Coefficient**	**Marginal Effect**	**Coefficient**	**Marginal Effect**	**Coefficient**	**Marginal Effect**
Taita	0.323	0.041	0.529*	0.079	0.472	0.073
	(0.299)	(0.047)	(0.313)	(0.065)	(0.292)	(0.061)
Work Type						
Seasonally Working (*Base - Not working/occasionally*)	0.064	0.006	0.042	0.004	0.075	0.006
	(0.085)	(0.009)	(0.087)	(0.009)	(0.080)	(0.009)
Permanently Working (*Base - Not working/occasionally*)	-0.120	-0.011*	-0.205**	-0.020**	-0.174**	-0.019**
	(0.073)	(0.007)	(0.087)	(0.008)	(0.079)	(0.008)
Education level						
Incomplete Primary (*Base – No education or Pre-Primary school*)	0.226*	0.023*	0.234*	0.025*	0.203*	0.024*
	(0.118)	(0.013)	(0.126)	(0.014)	(0.118)	(0.014)
Primary	0.197	0.021	0.201	0.022	0.173	0.018
	(0.127)	(0.015)	(0.137)	(0.016)	(0.128)	(0.016)
At least Secondary	0.143	0.015	0.085	0.009	0.065	0.004
	(0.133)	(0.015)	(0.161)	(0.017)	(0.149)	(0.017)
Higher	0.014	0.001	-0.199	-0.017	-0.239	-0.021
	(0.201)	(0.020)	(0.289)	(0.022)	(0.258)	(0.018)
Other Individual Characteristics						
First sex intercourse below 15 years of age	0.097	0.010	0.068	0.007	0.067	0.007
	(0.070)	(0.007)	(0.073)	(0.008)	(0.065)	(0.007)
Individual aged 15 to 35 years (*base – aged 36 – 49*)	0.116	0.011	0.154*	0.014*	0.119	0.013*
	(0.085)	(0.007)	(0.087)	(0.008)	(0.081)	(0.008)
Husband lives away from the family	-0.233*	-0.019**	-0.241*	-0.020**	-0.245**	-0.019**
	(0.123)	(0.009)	(0.131)	(0.009)	(0.123)	(0.010)
Married (*Base- Never married*)	0.341***	0.033***	0.433***	0.042***	0.338***	0.037***
	(0.083)	(0.008)	(0.105)	(0.011)	(0.096)	(0.009)
Just living with partner	0.375***	0.048**	0.530***	0.077**	0.408***	0.064**
	(0.136)	(0.022)	(0.155)	(0.031)	(0.143)	(0.026)
Widowed	0.891***	0.163***	0.996***	0.197***	0.808***	0.189***
	(0.141)	(0.039)	(0.153)	(0.047)	(0.201)	(0.044)
Divorced	0.670***	0.108**	0.793***	0.142**	0.654***	0.126**
	(0.211)	(0.049)	(0.212)	(0.057)	(0.228)	(0.053)

Table 2.4. ***continued***

Dependent Variable: Individual with HIV/AIDS

	Probit Model (1)		IVProbit Model (2)		Heckprobit Model (3)	
Variables	**Coefficient**	**Marginal Effect**	**Coefficient**	**Marginal Effect**	**Coefficient**	**Marginal Effect**
Separated	0.596***	0.090***	0.696***	0.115***	0.571***	0.105***
	(0.129)	(0.027)	(0.146)	(0.036)	(0.158)	(0.032)
No sex abstain	0.083	0.008	0.097	0.010	0.119*	0.010
	(0.065)	(0.006)	(0.068)	(0.007)	(0.062)	(0.007)
No condom use during sex	-0.114*	-0.011*	-0.102	-0.010	-0.105*	-0.012*
	(0.063)	(0.006)	(0.066)	(0.007)	(0.062)	(0.006)
Many sex partners	0.133**	0.013*	0.173**	0.018**	0.166***	0.016**
	(0.066)	(0.007)	(0.069)	(0.008)	(0.063)	(0.008)
Prostitution	-0.057	-0.006	-0.123	-0.013	-0.128	-0.009
	(0.110)	(0.012)	(0.116)	(0.014)	(0.108)	(0.013)
Male circumcision	-0.237*	-0.020**	-0.141	-0.013	-0.146	-0.015
	(0.134)	(0.010)	(0.137)	(0.011)	(0.124)	(0.011)
Constant	-2.633***		-2.502***		-1.925***	
	(0.200)		(0.202)		(0.450)	
/athrho			-0.144		-0.666*	
			(0.104)		(0.394)	
/lnsigma			-0.354***			
			(0.024)			
Rho			-0.143		-0.582	
			(0.102)		(0.260)	
Sigma			1.425			
			(0.034)			
Wald test of independent equation:			(/athrho= 0):	$\chi_1^2 = 1.93$ $\Pr > \chi^2 = 0.164$	(rho= 0):	$\chi_1^2 = 2.86$ $\Pr > \chi^2 = 0.091$

Table 2.5. Models of Wealth *(Model 2)* and Selection into HIV Testing *(Model 3)*

	Dependent Variables	
Instruments/Other Variables	**Wealth**	**Selection for HIV Testing**
Husband living away	0.154*	
	(0.081)	
Interview Interruption		-0.176*
		(0.105)
Occupation		
Professional, technical or Managerial or Clerical (*Base – Not working or Unknown*)	0.467	
	(0.301)	
Sales	-0.208	
	(0.259)	
Agricultural - self employed	-0.455*	
	(0.253)	
Household & Domestic	1.737***	
	(0.304)	
Services, Skilled & Unskilled Manual	-0.196	
	(0.269)	
Religion		
Catholic (*Base – Other or No religion*)		0.252
		(0.156)
Protestant or Seventh Day Adventist		0.254*
		(0.150)
Muslim		0.190
		(0.166)
Household Variables		
Wealth, Continuous		-0.030
		(0.046)
Wealth Continuous residuals		-0.004
		(0.048)
Gender of household head (*Base-Female*)	0.371***	0.017
	(0.090)	(0.066)
Household member gender (*Base-Female*)	0.120***	-0.027
	(0.044)	(0.074)
Polygamy	-0.079	
	(0.086)	
Household size	0.023	
	(0.020)	
Residence		
Urban (*Base- rural*)	1.426***	-0.361***
	(0.112)	(0.102)
Coast province (*Base - Nairobi*)	-0.213	0.603***
	(0.258)	(0.134)
Central province	-1.104***	-0.043
	(0.206)	(0.113)
Eastern province	-1.327***	0.124
	(0.184)	(0.132)
Nyanza province	-1.728***	0.648***
	(0.219)	(0.157)
Rift valley province	-1.429***	0.168
	(0.157)	(0.110)
Western province	-1.945***	0.450***
	(0.220)	(0.167)

Table 2.5. ***Continued***

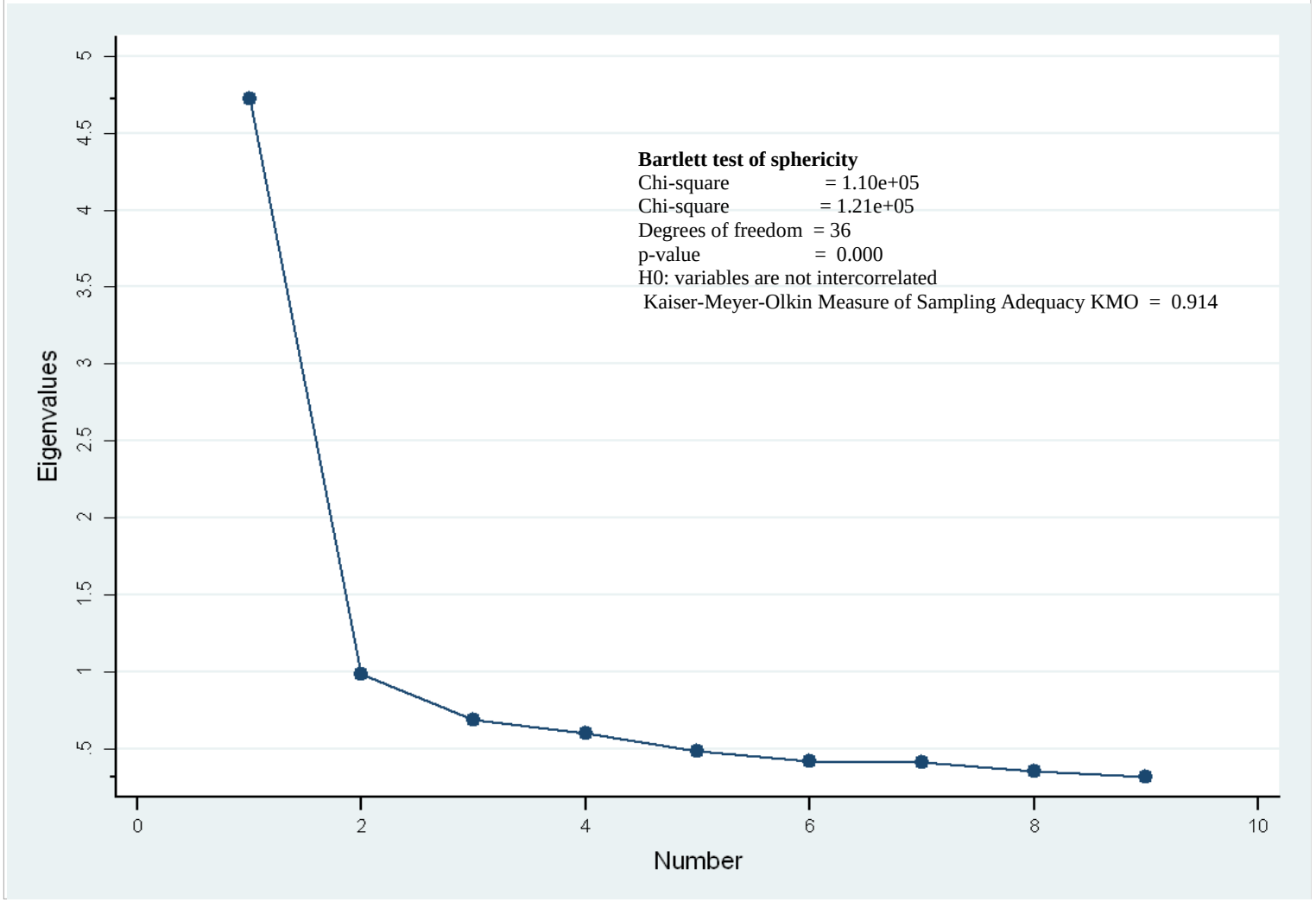
Other Variables		
	Wealth	**Selection for HIV Testing**
Other Individual Characteristics		
Tribe		
Kalenjin (*Base - Other*)	0.209	0.301**
	(0.160)	(0.126)
Kamba	-0.091	0.019
	(0.157)	(0.105)
Kikuyu	0.147	-0.066
	(0.174)	(0.100)
Kisii	-0.084	0.432***
	(0.213)	(0.164)
Luhya	0.116	0.194
	(0.199)	(0.128)
Luo	0.223	0.079
	(0.217)	(0.130)
Mijikenda/Swahili	-1.025***	-0.473***
	(0.262)	(0.157)
Taita	-1.274***	-0.077
	(0.313)	(0.218)
Work Type		
Seasonally Working (*Base - Not working/occasionally*)	0.307	-0.058
	(0.247)	(0.065)
Permanently Working (*Base - Not working/occasionally*)	0.507**	-0.001
	(0.245)	(0.054)
Education level		
Incomplete Primary (*Base – No education or Pre-Primary school*)	0.594***	-0.010
	(0.123)	(0.088)
Primary	0.676***	-0.056
	(0.137)	(0.097)
At least Secondary	1.331***	-0.077
	(0.155)	(0.116)
Higher	2.909***	0.022
	(0.233)	(0.198)
Other Individual Characteristics		
First sex intercourse below 15 years of age	0.048	-0.014
	(0.073)	(0.058)
Individual aged 15 to 35 years (*base – aged 36 – 49*)	-0.217***	0.010
	(0.077)	(0.056)
Husband lives away from the family	0.156	0.137
	(0.098)	(0.085)
Married (*Base- Never married*)	-0.578***	0.011
	(0.089)	(0.062)
Just living with partner	-0.820***	0.012
	(0.132)	(0.108)
Widowed	-0.382***	0.227*
	(0.134)	(0.134)
Divorced	-0.617**	0.020
	(0.272)	(0.228)
Separated	-0.531***	0.054
	(0.161)	(0.113)
No sex abstain	0.017	-0.092*
	(0.070)	(0.048)

Table 2.5. ***Continued***

Other Variables		
	Wealth	**Selection for HIV Testing**
No condom use during sex	-0.141**	0.004
	(0.066)	(0.047)
Many sex partners	-0.171**	-0.065
	(0.071)	(0.048)
Prostitution	0.341***	0.138*
	(0.103)	(0.072)
Male circumcision	-0.264***	-0.238***
	(0.079)	(0.053)

(*) Marginal effects are for discrete change of dummy variable from 0 to 1
* significant at 10%; ** significant at 5%; *** significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)

Figure 2.1. **Scree Plot of eigenvalues after PCA for household assets**



Appendix 2.A: Variable Glossary

Dependent Variables	
Any household Individual with HIV/AIDS	Dummy = 1 if individual was HIV positive given was selected and agreed for test
Household member agree for HIV test *(for selection equation)*	Dummy = 1 if the individual had agreed to be tested for HIV given was selected
Independent Variables	**Description**
Household characteristics	
Household head	Dummy = 1 if the household head is a male and 0 if female
Household member gender	Dummy = 1 if household member is male and 0 if female
Polygamy	Dummy = 1 if woman is married with other wife (ves) and 0 if only wife or single.
Wealth Index	
Poorest household Poorer household Middle household Richer household Richest household	Dummy = 1 if the individual is in the classified household wealth named and 0 otherwise
Wealth, Continuous	Predicted score from the first principle component from PCA for household assets
Wealth Continuous residuals	Residuals after regressing Wealth (Continuous) variable on selected explanatory variables.
Residence/Region	
Urban (*Base Rural*)	Dummy = 1 if the individual is in the household residing in urban area and 0 if Rural areas
Nairobi *(Base)* Central Coast Eastern Nyanza Rift Valley Western North Eastern	Dummy = 1 if the individual is from the household residing in named the Province and 0 otherwise
Individual characteristics	
Tribe	
Kalenjin	Dummy = 1 if the individual is from Kalenjin tribe and 0 if other tribe
Kamba	Dummy = 1 if the individual is from Kamba tribe and 0 if other tribe
Kikuyu	Dummy = 1 if the individual is from Kikuyu tribe and 0 if other tribe
Kisii	Dummy = 1 if the individual is from Kisii tribe and 0 if other tribe

Luhya	Dummy = 1 if the individual is from Luhya tribe and 0 if other tribe
Luo	Dummy = 1 if the individual is from Luo tribe and 0 if other tribe
Masai	Dummy = 1 if the individual is from Masai tribe and 0 if other tribe
Meru	Dummy = 1 if the individual is from Meru tribe and 0 if other tribe
Mijikenda/Swahili	Dummy = 1 if the individual is from Mijikenda/Swahili tribe and 0 if other tribe
Somali	Dummy = 1 if the individual is from Somali tribe and 0 if other tribe
Taita	Dummy = 1 if the individual is from Taita tribe and 0 if other tribe
Turkana	Dummy = 1 if the individual is from Turkana tribe and 0 if other tribe
Work type	
Seasonally Working	Dummy = 1 if an individual work is seasonal and 0 if Otherwise
Permanently Working	Dummy = 1 if an individual work is permanent and 0 Otherwise
Not Working	Dummy = 1 if an individual is Not working or works occasionally and 0 Otherwise
Educational Level	
No education	Dummy = 1 if individual has pre Primary or no education and 0 otherwise
In complete Primary (*Base – No education or Pre-Primary school*)	Dummy = 1 if individual's education level is incomplete Primary and 0 otherwise
Primary (*Base – No education or Pre-Primary school*)	Dummy = 1 if individual's education level is Primary and 0 otherwise
At least Secondary	Dummy = 1 if individual's education level is at least Secondary and 0 otherwise
Higher	Dummy = 1 if individual's education level is Higher and 0 otherwise
Marital Status	
Never Married	Dummy = 1 if individual has never married and 0 otherwise
Married	Dummy = 1 if individual is married and 0 otherwise
Just living with partner	Dummy = 1 if individual just lives together with the partner and 0 otherwise
Windowed	Dummy = 1 if individual is a window and 0 otherwise
Divorced	Dummy = 1 if individual is divorced and 0 otherwise
Separated	Dummy = 1 if individual is separated from the partner and 0 otherwise
Other Individual's Characteristics	
No sex abstain	Dummy = 1 if individual does not abstain from sex and 0 if abstains
No condom use during sex	Dummy = 1 if individual does not use condom during sex and 0 if uses condoms
Many sex partners	Dummy = 1 if individual has many sex partners and 0 if has one sex partner
Prostitution	Dummy = 1 if individual do prostitution and 0 if not
No injection avoidance	Dummy = 1if individual does not avoid injections and 0 if avoids

No HIV test before sex relation	Dummy = 1 if individual/partner does not test for HIV before any sex relationship and 0 if is tested
Male circumcision	Dummy = 1 if a male individual is circumcised and 0 if not
Instruments	
Household size	Number of household members.
Aware of HIV	Dummy = 1 if an individual had knowledge or aware of HIV disease
Interview Interruption	Dummy = 1 if respondent's was interrupted by any adult from same household
Religion	
Catholic	Dummy = 1 if the individual in the household belongs to Catholic religion and 0 otherwise
Protestant and other Christianity	Dummy = 1 if the individual household belongs to Protestant/Other Christian religion and 0 otherwise
Muslim	Dummy = 1 if the individual household belongs to Islamic religion and 0 otherwise
No Religion	Dummy = 1 if the household belong to no religion and 0 otherwise
Other religion	Dummy = 1 if the individual household belong to Other religion and 0 otherwise
Occupation	
Professional, technical or Managerial	Dummy = 1 if an individual's occupation is Professional, technical or Managerial and 0 otherwise
Clerical	Dummy = 1 if an individual's occupation is Clerical and 0 otherwise
Sales	Dummy = 1 if an individual's occupation is Sales and 0 otherwise
Agricultural - self employed	Dummy = 1 if an individual's occupation is Agricultural self employed and 0 otherwise
Household & Domestic	Dummy = 1 if an individual's occupation is Household and Domestic work and 0 otherwise
Services, Skilled & unskilled Manual	Dummy = 1 if an individual's occupation is services or skilled or unskilled Manual and 0otherwise
Unknown	Dummy = 1 if individual's occupation is unknown or Not working and 0 otherwise
λ	Inverse Mills Ratio: Selectivity bias correction factor computed from the estimated household member testing HIV/AIDS positive

CHAPTER 3

IMPACT OF HOUSEHOLD CHARACTERISTICS, SOCIO-CULTURAL FACTORS AND HIV/AIDS ON CHILD SCHOOLING

3.1 Introduction

In this chapter we examine the issues surrounding a child acquiring formal education. Specifically, the study deals with children's education outcomes, examining school attendance, school attainment, and rates of grade progression. Education brings significant economic benefits to individuals and communities. For individuals, these include a higher standard of living, greater economic independence and also enhanced opportunities for personal development (travel, car, computer, house, etc.), together with other factors such as child health and nutrition, which improve community welfare overall.

Lack of education excludes countries and individuals from opportunities to improve their living standards and quality of life. Education also makes democracy possible by creating a demand for individuals or a party to be heard and to enhance people's ability to influence decisions which affect their lives and the country at large. For instance, when educated individuals demand that governments, schools, and health clinics be made more accountable, the entire community benefits since the intended project is done with less corruption and in a proper way. All these reasons and probably much more, drive the need of education. It is in the country's best interests to extend opportunities

for at least basic education. As a result the public provision of education is seen as a requirement for social and economic progress (Galbraith, 1996). Education enhances economic growth by externalities generated by the education of members of a society. It is on this basis that education is vital and must be encouraged for future productivity of an economy.

There are of course economic benefits from sending children to school, primarily in the form of a wage premium - the expected rise in earnings potential as a result of greater educational attainment. Further, in developing countries in the absence of well-developed financial and insurance markets, children also act as an informal source of old-age security for their elderly parents (Ray, 1998; Nugent, 1985; Nugent & Gillaspy, 1983). The higher the education level attained by the child, the greater is their earnings potential as adults, thus increasing the likelihood of higher potential transfers from children in their parent's retirement. At the household level, this provides parents with a motive for investment in their children's schooling. One strand of literature, such as early studies by Becker and Lewis (1973), examine this issue in the context of a household production model where fertility and child schooling decisions are made jointly. The essential idea here is that as households are resource constrained, there is a negative relationship between family size and schooling enrolment in the household. This has been termed in the literature as the "quantity-quality" trade off. This approach has primarily been used to identify policies that are likely to result in reductions in fertility and increases in educational investment at the household level.[22]

[22] There are a number of studies using data from developed countries (US particularly) that find that children with fewer siblings obtain more schooling than those with more siblings and this negative relationship persists even when family socioeconomic characteristics are controlled for. Evidence from developing countries is however quite mixed. Our results do not support the quantity-quality trade-off in

The lack of human capital is the most serious constraint in developing countries. Thus, investment in human capital through education is universally recognised as an essential component of economic development. While education endows individuals with the means to enhance their skills, knowledge, health and productivity, it also enhances the economy's ability to develop and adopt new technology (Ray, 1998). Hence, increasing education levels is an important policy concern in most countries (Maitra, 2003). Families invest in children's education for many reasons, among them the expectation that education will increase the child's future earnings. Governments also invest in education in order to raise the skill level of the labour force and, hence, to increase worker productivity and income in society at large (Deaton, 1997).

3.2 Education history in Kenya

Before the coming of Western education, traditional African and Islamic education systems were practiced in Kenya. Education was based on life experiences, moral values and skills necessary for life. African indigenous education conserved and transmitted knowledge and wisdom from one generation to another (Mugai, 2002).

3.2.1 Education System in Kenya

The introduction of Western education in the mid 19th century by the Christian missionaries disrupted the balanced indigenous economy without giving quick returns. The belief that education would change one's livelihood brought such a keen hunger for education that after independence every parent in most of the communities in Kenya

schooling and we explain our results in the form of existence of economies of scale in educating children and in the development of schooling norms within the household.

wanted their child to go to school (Olson, 1972). However, academic learning was not the sole agenda of the missionaries and the mission schools called “village schools” did not offer education beyond the fourth year of elementary education (Bogonko, 1992). The white settlers needed educated labour in the form of masons, carpenters and other trades and thus did not support literacy learning. The most concentrated educational efforts were in those parts of Kenya adjacent to Kisumu and Nairobi; and it is in these areas that schooling was most rapidly and thoroughly diffused (Olson, 1972).

As in any other colonial era, education in Kenya was racially stratified, with separate schools for Whites, Asians, Arabs and Africans. Up to Kenyan independence in 1963, there was great disparity in education levels between the different races and also between males and females since females were mostly not participating in the education system as the men were being trained to work for the colonial government (Bogonko, 1992).

After independence in 1963, the government of Kenya embarked on a massive expansion of the education system to make education accessible to all people who could not get it during the colonial era, although this was not achieved in an even manner. The Kenyan curriculum was changed to reflect Kenyan cultural values but still centred on passing national examinations. Education was seen as a vehicle to;

i. train more human resources to enhance economic development;
ii. acquire a suitable basic foundation for further education, training and the world of work;

iii. develop awareness and understanding of the immediate environment and foster positive attitudes towards other countries and towards the international community;

iv. develop a strong whole person, including the physical, mental and spiritual capacities; and;

v. develop desirable social standards and attitudes.

The Kenyan government built more "Harambee"[23] schools and boarding school mushroomed in every location in the country. Because of problems in financing, however, only a few of these self-help schools had managed to open fourth form classes. The government, in an attempt to accommodate these rapidly emerging schools, initiated a new examination at the end of form two and only those who did well were admitted into form three of the government schools. Suffice it to say that in spite of this self-help effort, government schools have remained the prime source of potential elites in Kenya (Olson, 1972; Eshiwani, 1993; Eisemon, 1988; Eisemon & Nyamete, 1988)

From 1963 to 1983 primary education was for seven years but in 1984 the government implemented a system of eight years in primary education (Standards), a four year secondary (Forms) and a four year-year university (currently referred to a 8-4-4 education system). The language of instruction is English throughout the school system (in some areas indigenous languages are used in first or second grades). In addition to government schools, there are a number of private schools, many of which serve Asian and European communities. The 8-4-4 system was created to help those students who

[23] *Harambee* is a Swahili word meaning "self-help" and is commonly used to mean come together and combine our resource efforts in terms of labour or funds to enable the contributions or capital for the community project or help an individual, eg building schools, hospitals, clearing community member bills or funding especially higher education (Chieni, 1999).

are not planning on furthering their education after secondary school acquire skills that will help them find employment (Eshiwani, 1993; Eisemon , 1988; Eisemon & Nyamete, 1988; Ministry of Education Science and Technology [MEST], 1996, 1998, 2001, 2004).

In 1963 the Kenyan government promised free primary education but this did not take effect until 2003[24]. Before this time, due to their inability to pay fees, many children were locked out of school. The official entry age to primary school is 6. (Eshiwani, 1993; Eisemon, 1988; Eisemon & Nyamete, 1988; MEST, 1996, 1998, 2004).

Kenya's thirst for more knowledge has exceeded expectations. After the first university was established in 1970, five other public universities have been created. Several private universities have also been established. In addition, a number of post-secondary institutions offer training at diploma and certificate levels (Eshiwani, 1993; MEST, 1996, 1998, 2004).

This study analyses the issues surrounding children's primary schooling and grade attainment using the data of KDHS for 2003. The roles of individual and household characteristics are investigated. The remainder of this chapter is arranged as follows: the next section gives the literature review of relevant previous research work, followed by an overview of economic growth in relation to human capital, taking education as a main component of human capital. After this, the next section introduces the econometric methodology and data used, then definition of variables are followed by the econometric specification. Results come after this, followed by the discussion and conclusion with policy implications.

[24]Schools became overcrowded and faced issues like less teachers and poorer teaching facilities. Some parents who could afford to pay more tuition fee sent their children to private schools.

3.3 Literature on Child schooling

The theory of economics of education was initiated by Schultz (1961), Mushkin (1962), Becker (1964), Fuchs (1966) and Mincer (1974) where they considered education as an investment and not a consumption activity. For example, in his first edition, Becker considered education and training as the most important investments in human capital. However, shortly after this association, other researchers including Arrow (1973), Spence (1973) and Stiglitz (1975) came up with a different idea that education was acting as a private signal to the employer and was not a social value.

Becker (1975) established that education is an investment and it adds to our human capital just as other investments add to physical capital. He further pointed out that investment in education and training creates the human capital (basically the skills and abilities) that is a vital element in assuring economic growth and individual advancement and reducing inequality. Becker (1975, 1994) shows that as a result of education, some of the returns to this investment can be measured while others cannot (see also Mincer, 1974). More recently, some economists considered education's role in endogenous growth (Romer, 1986, 1990). Lucas (1988) pointed out that an individual's human capital enhances the productivity of other factors of production, such as physical capital and the human capital of others, through channels that are not internalised by individual families or firms.

After 1990, research has focused on identifying the education externalities in economic growth and quantifying non-market effects (Venniker, 2001). For example, Benhabib and Spiegel (1994) in their study gave a negative assessment of the relationship between

human capital and productivity, let alone the existence of positive externalities. Also, Acemoglu and Angrist (2001) indicated that for education to raise income there must be human-capital externalities. In contrast, Card (1999) pointed out that the increase in individual earnings which resulted from an additional year of schooling was about 6 – 10%. Human-capital externalities are important for education policy as well as for explaining cross-country income differences (Acemoglu & Angrist, 2001). Many other research papers have worked on human capital externalities with reference to education (see Manda, Mwabu & Kimenyi, 2004; Halfdanarson, Heuermann & Sudekum, 2008; Massimiliano & Leombruni, 2009; Bakis, *et al.*, 2009).

Some of the literature has focused on educational outcomes, treating fertility as an exogenous variable. See for example studies by Gertler and Glewwe (1992), Butcher and Case (1994), Glewwe and Jacoby (1994), Jacoby (1994), Kaestner (1997) and Psacharopoulos (1997). Mostly this approach has been used to examine policies that might be used to maximise the effectiveness of public spending on education, including issues relating to the ability and willingness of households to bear some of the costs of provision of educational services.

In this study we are going to examine school enrolment and education attainment treating the number of siblings as exogenous and estimating reduced form regressions of demand for education. Any analysis of the effectiveness of schooling should ideally account for the final level of schooling that an individual attains and relate this to information on the environment in which this individual grew up. However most survey data sets from developing countries are non-retrospective and provide very little information on the environment in which the adult grew up. There is therefore very little

information on the factors that typically affect educational attainment. Due to this fact, most studies are carried out with the child's school engagement as the focus. Chernichovsky (1985) and Maitra (2003) point out that studies using children as the unit of observation allows one to use available information on parental, household and community characteristics and hence information on the environment in which schooling decisions are being made.

The economic and social returns to investment in basic education in an agricultural country like Kenya have been higher from the time of independence than other forms of educational investment. Compared to other countries in Sub-Saharan Africa with similar GDP per capita, Kenya spends considerable more on education in relation to total Government expenditure and Gross National Product (Abagi, *et al.,* 2000). In spite of huge investment in the 8-4-4 system of education by the government, parents, non-governmental organisations, and donors, enrolment at various levels of education is characterised by regional and gender disparities and declining gross enrolment ratios. Likewise, the quality and relevancy of education at all levels had been questioned and the education system has experienced high wastage as a result of high repetition and drop-out rates (Abagi, 1997a, 1997b; Ministry of Education [MoE], Human Resources Development [HRD], World Bank [WB] and Ministry of Health [MOH], 2001; Ministry of Education Science and Technology, 1996).

In their work, using 1996 data, Abagi and Odipo (1997) found that drop-out and repetition rates were higher in upper classes (Standards 5 to 8), compared to lower classes. The study showed that every year, about 10% of pupils from each class failed to move on to the next and this resulted in high cumulative loss, which was experienced by

Standard 8. As an example, in 1993, there were 472.5 and 384.2 thousand boys and girls enrolled in Standard 1 respectively. Four years later, only 372.9 and 364.2 thousand boys and girls were enrolled in Standard 4 giving a dropout rate of about 21% and 5% for boys and girls, respectively. The annual average drop outs rates by provinces revealed that North-Eastern Province had the highest drop-out rate of 9.4% followed by Western Province 8.0%, Nyanza 6.5%, Rift Valley 5.8% Eastern 5.6% and Central Province 2.2% and the national average drop-out rate was 5.4% (5.5% for boys and 5.3% for girls).

In another study, Abagi, *et al.* (2000) indicated declining enrolment and participation rates over time. They found that of the total enrolment, over a ten year period, only 46% of pupils managed to complete the primary education cycle (35% of girls and of 55% boys) and in 1999 the transition rate from primary to secondary was only 44.8% (43.1% for girls and 46.1% for boys). They also noted the widening gender and regional disparities particularly in the Arid and Semi-Arid Lands (ASAL) where the gross enrolment ratio at the national level in 1998 showed less gender disparity (49% of girls and 51% of boys). They found that the disparity was worse for North Eastern Province where the ratio was 16.8% for girls and 32.0% for boys (total 24.8%), followed by the Coast Province where boys had a ratio of 79.6% and girls 66.9% (total 73.3%) while in Nairobi Province the ratio was 61.6% for boys and 52.8% for girls (56.9%).

Bedi, *et al.* (2004) note that since independence, the Kenyan educational system has witnessed several changes in structure and in curriculum. From data collected in 1997, 44% of the working-age population had not completed primary school while 21% had attained at least 8 years of schooling (completed primary school), about 17% had begun

but not completed lower secondary education (forms 1 and 2) while 13.7% had completed (see also Kimalu *et al.,* 2001).

In their paper, Abagi and Odipo (1997) examined issues of efficiency in primary education in Kenya. School completion rates had remained below 50% in the last five years. They also noted that teaching-learning time was not utilised efficiently in primary schools. They attributed these problems to education policies and management processes (mis-allocation of resources to educational levels), school based factors (teachers attitudes, time utilisation, school environment), and household based factors (poverty, socio-cultural factors, and gender issues). In their view, education policy needed to be restructured to reduce drop-outs and enhance completion rates.

Research shows that one of the main household and community based factors affecting child schooling is the level of poverty in Kenya which discourages parents from investing in their children's education.[25] As a result of the introduction of the cost-sharing policy in 1988, parents are expected to meet 95% of their children's education costs and this limits many Kenyan children's access to education (Abagi, 1997a; MEST, 1996). Increasing poverty means child labour becomes crucial for family survival. Most child labour involves domestic activities, agriculture, or petty trade in rural and urban Kenya.[26]

Socio-cultural and religious factors such as initiation ceremonies and gender socialisation are additional factors responsible for pupils' failure to complete primary

[25] The 1997 Economic Survey indicated that 46.8% of Kenyans lived below the poverty line.
[26] Parents have continued to send their children, particularly daughters, into the labour market—mainly as domestic workers in urban centres. In case of boys mostly from the coastal region and in rich agricultural areas abandon school in order to earn money as beach-boys and tea or coffee pickers, respectively.

education. In areas where traditional circumcision is still practised, some pupils are pulled out of school to participate in initiation ceremonies. Some circumcised boys treat women as inferior and cannot be taught anything by women. Similarly, some initiated girls feel that they are now grown up women who should get married. This is because, in some communities, girls or boys expected to get married immediately after they have been initiated. Pressure is therefore put on them to leave school and meet traditional expectations (Abagi & Odipo, 1997).

In our empirical analysis we will specifically look at the impact of HIV/AIDS on children's education. This disease can be seen as barrier to development gains in many countries, not only Kenya, leaving populations more vulnerable to poverty, malnutrition, and ill health. One of the consequences is the number of orphaned children, particularly in sub-Saharan Africa (UNICEF 2003; UNAIDS, UNICEF & USAID, 2004). The burden of caring for the sick parent is often taken up by the children and many are forced to drop out of school to take up the roles of their parents (see Case, Paxson & Ableidinger, 2004). After the death of the parents, household resources are reduced due to the loss of income, which in turn affects health care and the nutritional status of children, as well as their educational status. What often follows after the death of parents is changes in living arrangements, displacement, and fewer resources for schooling, health and food for children.

Bedi, e*t al.* (2004) showed that the correlations between HIV prevalence rates and enrolment rates did not suggest any relationship. In fact for some years the relationship between the two rates was positive. Temporal patterns also suggest that there is no relationship between change in HIV prevalence rates and change in enrolment rates.

However, they had not controlled for the urban location of households nor for any other individual, family or regional characteristics. In our study we have incorporated the variable for HIV/AIDS in our model to examine its effect on child's schooling outcome and we have controlled for urban/rural location and also for individual, family and child's gender characteristics.

3.4 Theoretical Framework

The household decision model is the most widely used method of analysis in the child schooling literature. Decisions on whether or not to send a child to school are typically made by weighing the costs and benefits of schooling, depending on other historical characteristics which include household educational background, level of wealth/poverty, traditional customs and social cultures (Deaton, 1997). For poor resource-constrained households particularly, the cost of sending children to school including school fees, school uniforms, books, transport etc. can be quite significant. Furthermore, there is also an opportunity cost attached to school attendance in the form of foregone labour contributions. These costs can be very high for poor households residing in rural areas, particularly where children contribute substantially by working in the market for a wage or by contributing to household chores such as farm cultivation, cattle keeping, caring for younger siblings, fetching water and firewood, cleaning.

Wilson (2001) has given a structural model where family background and school quality are allowed to affect the returns to education as well as the utility costs and benefits of acquiring education. In her model, it is assumed that the individuals make their education decisions based on the expected returns and costs involved. Similar models

were presented by Checchi (2006) when constructing a model of education as an investment in human capital. Checchi's approach was more based on Ben-Porath (1967), and produces a closed solution for the optimal path of investment.

In our model we start by dividing the life of an individual into two periods; during youth (period 1) and adulthood (in period 2). In order to increase his or her stock of human capital, H by schooling in a certain period, an individual (specifically a child) has to devote a part S of his or her time attending school. Investment in education depends on many factors including the school quality characteristics Q (eg. School type and facilities, teacher experience, etc), child characteristics C (eg. child's ability, health, age), household characteristics R (eg. wealth, size etc), other educational inputs prices E (school fees, books, school distance from home etc). The stock of human capital is intended to have an impact on future labour market rewards. Putting these into equation form, human capital evolves according to

$$H_2 = H_1(1-\delta) + \Delta H \tag{3.1}$$

where the acquired human capital is expected to depreciate with time at a rate of $\delta \geq 0$ and

$$\Delta H = H(S_1, Q_1, R_1, C_1, E_1, H_1) \tag{3.2}$$

we can therefore construct the individual's discounted lifelong earnings:

$$N = W_1(H_1)(1-S_1) - \gamma_1 S_1 + \frac{W_2(H_2)(1-S_2) - \gamma_2 S_2}{1+\rho} \tag{3.3}$$

where γ is the direct cost of schooling which include the school tuition fees, books purchases, transport cost and any other schooling cost, and ρ is the intertemporal discount rate.

A child in school forgoes income in favour of schooling which contributes to their stock of human capital. In equation (3.3) this forgone income is captured by $W_1(H_1)S_1$.

To obtain the demand for education from the optimal choice of investment in human capital, we then maximize the utility (3.3) above with respect to S_1 and S_2 as follows:

$$\max_{S_1,S_2} N = \max_{S_1,S_2}\left[W_1(H_1)(1-S_1) - \gamma_1 S_1 + \frac{W_2(H_2)(1-S_2) - \gamma_2 S_2}{1+\rho}\right] \tag{3.4}$$

This function can be minimised by taking first order condition with respect to each of the two decision variables, S_1 and S_2. Obtaining the first order condition with respect to S_2, shows that $\frac{\partial N}{\partial S_2} < 0$ so the optimal $S_2 = S_2^* = 0$ meaning that there is no value in investing on education in the second period (adulthood). This is because S_2 does not contribute to human capital in period 2, but time spent in education reduces time available for work, and hence labour income.

Concentrating out $S_2^* = 0$, then equation (3.3) becomes

$$N = W_1(H_1)(1-S_1) - \gamma_1 S_1 + \frac{W_2(H_2)}{1+\rho} \tag{3.5}$$

We next maximize the utility (3.5) above with respect to S_1 to give

$$\frac{\partial N}{\partial S_1} = -W_1(H_1) - \gamma_1 + \frac{1}{1+\rho}\frac{\partial W_2(H_2)}{\partial H_2}\frac{\partial H_2}{\partial S_1}$$
$$= 0 \text{ for maximum} \qquad (3.6)$$

From equation (3.1), $\frac{\partial H_2}{\partial S_1} = \frac{\partial \Delta H}{\partial S_1}$

This implies that the optimal S_1 is chosen so that

$$\frac{1}{1-\rho}\frac{\partial W_2(H_2)}{\partial H_2}\frac{\partial \Delta H}{\partial S_1} = W_1(H_1) + \gamma_1 \qquad (3.7)$$

The right hand side of (3.7) is the cost of investing in education in period 1: the direct cost γ_1 and the cost in lost earning during education $W_1(H_1)$. The left hand side is the returns in period 2 from investing in schooling in period 1. This occurs via increased human capital $\frac{\partial \Delta H}{\partial S_1}$, that in turn feeds into higher earnings via $\frac{\partial W_2(H_2)}{\partial H_2}$.

We assume $\frac{\partial^2 \Delta H}{\partial S_1^2} < 0$, diminishing returns to schooling in building human capital and we also assume $\frac{\partial^2 W_2(H_2)}{\partial H_2^2} < 0$, diminishing returns to human capital. With both these diminishing returns, there will come a point where the costs of schooling outweigh the benefits. The equation to be estimated thus becomes:

$$S_1^* = S\left(Q_1, R_1, C_1, E_1, H_1, \gamma_1\right) \tag{3.8}$$

The partial derivatives of equation (3.8) give the marginal utilities. From this optimal demand equation, it tells us that more the able (given by unobservable ability, C) individual tends to demand more education so $\frac{\partial S_1^*}{\partial C_1} > 0$. Similarly those attending a better school (characterised by Q) or with other better educational inputs (given by E) will get higher educational returns since they accumulate more human capital per unit time. Therefore, the demand for education will be higher if the future expected gain is higher. Dehejia and Gatti (2002) give the evidence that the prohibition of child labour and subordinating the salary of children encourages child school attendance. In case of high cost (γ_1) of schooling, then the household becomes reluctant or unable to educate the child.

In the continuous-time case, Ben-Porath (1967) and Checchi (2006) have given a detailed model showing that the optimal investment path requires concentrating schooling in the initial part of a lifetime even if in practice, human capital accumulation continues after full-time education. In other words, the stock of human capital will increase reaching a maximum point and then will start depreciating, producing a decline in human capital.

3.5 Econometric Methodology and Data Issues

3.5.1 The Methodological Structure

In this section the focus is on the factors which have an impact on a child's schooling. In addition, we will give an overview of econometric methodologies used in previous research and on the basis of this consideration, we look into available data and econometric issues to structure the appropriate econometric methodology for this study.

3.5.1.1 Relevant literature on modelling child schooling

This is one area which has complications due to methodology and limitations of suitable data, especially when looking into the household characteristics which influence a child's education. For instance, when we want to model the highest level of schooling attained one would encounter complications. Maitra (2003) points out that a large mass point frequently characterises most data used for school attainment from developing countries at zero years of education. Similar spikes occur at primary and secondary school completion levels, where household and school resources and entrance requirements often impede progress to the next level. This leads to inappropriate analysis using Ordinary Least Squares (OLS) estimation as many studies have used, for example Handa (1996), and Parish and Wills (1993).

Due to this problem, the most common approach used in the literature is the use of ordered probit and logit models to estimate the highest grade attained, for example a study by Tansel (1997), and Dreze and Kingdon (2001). However, this has never resolved the problem completely. Only after ignoring the censoring in the data arising

from the fact that some children are enrolled in school at the time of the survey, would it be appropriate to use the ordered probit or logit model.

Turning to the analysis of the highest grade attained, the analysis is somewhat complicated by the problem of discreteness of the completed years of schooling, the problem of probability events (spikes) and right censoring. A censored ordered probit model could be suitable for estimation purposes (Maitra, 2003). This is in part because some children will start schooling late[27] (i.e older than the required age for initial school enrolment) and therefore, the age groups and education level attained do not match perfectly, hence the final complete level of education for the members of the cohort may not be necessarily known at the time of the survey. Therefore, those children who are still in school will be regarded as being censored (Cornwell, e*t al.*, 2005 and Maitra, 2003).

In this study when modelling attainment we have restricted the sample to children aged 15 - 18, as children in this age group should normally have completed primary schooling. This eliminates most but probably not all of the censoring effect. Since this censoring could be related to the individual, the household level and community characteristics that are used as explanatory variables, this may bias the estimated coefficients in the model used.

When current school enrolment is the dependent variable, this censoring issue is not a problem. However, there might be other problems associated with this particular variable. The estimation results will be biased if either the age at which children start schooling or the extent to which they repeat grades is related to the explanatory variables

[27] Some children start schooling earlier or jump a grade or class due to their higher ability.

that are used. For example, suppose that girls and boys have identical education attainment but most girls start school late or repeat more grades. In this case, at any point in time we expect to see more girls than boys enrolled in school, giving the false impression that girls receive more education than boys (see Cornwell, *et al.* (2005) and Anh *et al.* (1998). While we acknowledge these issues, it should be noted that the existing data does not allow us to correct for these potential problems.

These models are further complicated by the explanatory variables which include an ordinal endogenous variable, the variable for household wealth. Relatively little attention has been paid to models with an ordinal endogenous independent variable, including the problem of obtaining the correct standard errors (for more detail see Chapter 2, section 2.4.2).

In the work of Anderson and Lam (2003) and Anderson (2005) on grade progression, the dependent variable used has a known lower bound of 0 for those who have passed no grades or had no schooling and an upper bound of 1 for those progressing at a rate of one grade per year. In practice, the progression rates can be outside the defined theoretical bounds, possibly due to children beginning school at different ages and grade acceleration. If we truncate the data to the theoretical bounds and not include the potentially misleading results from the raw data, a double-sided tobit model would be a choice for a ratio model of this nature. There is one further limitation of the ratio-based model of Anderson and Lam (2003) and Anderson (2005). Since the years of education take only integer values, the ratio of integer values used by Anderson and Lam (2003) and Anderson (2005) to measure grade progression can only take a limited number of values. The ratio variable would also increase in variance with age. Neither the OLS nor

Tobit estimates account for either of these characteristics. Cornwell *et al*. (2005) also pointed out this problem and they use a different approach involving a binomial regression model.

In each regression model set we include all the children in the sample and as a result of this, some households contribute multiple children to the sample, hence used in estimation. Since unmeasured determinants of schooling are likely to be correlated within families (or households), the estimated standard errors are expected to be biased downwards and therefore, we allow for this correlation by adjusting the standard errors for clustering on households.

3.5.2 Definitions of Variables for this study

3.5.2.1 Dependent variable: School attendance

The current school enrolment (School Attendance) is a dichotomous indicator and is estimated using a probit model with an individual child as a unit of observation:

$$School\ Attendance = \begin{cases} 1 \text{ if the child is currently enrolled in school} \\ 0 \text{ otherwise} \end{cases} \qquad (3.9)$$

Current enrolment is analysed separately for children in the age groups of 6 - 14 and 15 – 18 years old. It is expected that different magnitude of effects could affect the decision to attend school for the two age groups. As mentioned earlier under the Kenyan education system these are the specified ages for primary and secondary schools.

3.5.2.2 Dependent variable: School grade attainment

The second dependent variable to be used in this study is grade attainment (School Grade Attainment), which is the highest grade achieved by the child in primary school level. The analysis of the highest grade attained is somehow complicated by the problem of discreteness of the completed years of schooling, the problem of probability events (spikes) and right censoring. To construct this model we have restricted the sample to children aged 15 - 18, as children in this age group should have completed primary schooling under normal progression rates. This eliminates some but not all of the censoring effect. To examine the grade attainment in this study we estimate the model ignoring the censoring issue and therefore the coefficients should be interpreted conditional on the assumption that any bias due to failure to take censoring into account, is minimal. Grade attainment is organised into three hierarchical categories as follows:

$$School\,Grade\,Attainment = \begin{cases} 0 \text{ if child was never enrolled to school} \\ 1 \text{ if child completed some but not all primary school} \\ 2 \text{ if child completed primary school} \end{cases} \quad (3.10)$$

3.5.2.3 Dependent variable: Grade Progression

The rate of grade progression of the child schooling is another important factor. The Kenyan education system requires a student to satisfactorily pass the current grade in order to go to the next class or grade. It is very common for children to repeat grades during the course of their schooling.

To define the dependent variable for the model of grade progression, we obtain the ratio of the actual grades completed to the number that the child should have completed given their age. We consider individuals aged 6-18 years old who are currently enrolled and attending primary or secondary school and not completed the final year of secondary school. The dependent variable "grade progression rate" is defined as follows:

$$Grade\ rate_i = \frac{\text{Current child grade}_i}{\text{Child's age}_i\ -\ 5\ \text{years}} \tag{3.11}$$

While our interest is in grade progression, where a child enrols in grade one late, she or he will be observed as having a slower progression rate compared to the child who enrolled at the required age. This would be problematic for interpretation if late commencement of schooling is correlated with the explanatory variables in our model used to estimate grade repetition. This is a point to be considered when interpreting and discussing the results. We need to note that estimating the grade rates using OLS would be fine because there is no maximum value for the grade rate of a child.

3.5.2.4 Choice of independent variables

To model the influence of family or household characteristics and other factors which affect a child's schooling, we will now discuss the explanatory variables used in the model. The variables will include individual and household characteristics (age of the child, child's gender, household head, household size, household dwelling characteristics, household wealth, parents marital status, ethnicity, religion, etc) and demographic and geographical factors.

The Individual (child) characteristics include the child's age and gender. In previous research on African schooling of children, Morduch (2000) argues that the gender composition of the household has an important influence on intra-household resource allocation of schooling and health resources, particularly if the child comes from a poor, resource-constrained household. This will imply that the educational attainment of children depends not only on their own gender but also on whether their siblings are male or female. The presence of other siblings in the same household can be seen as rivals in a competition for greater access to household resources.

To capture this effect in our study, the number, age and gender of children is grouped in the following, manner:

number of male and female children in the age group 0 - 5 (pre-primary school age);

number of male and female children in the age group 6 - 14 (primary school age; and

number of male and female siblings in the age group 15 - 18 (secondary school age).

To examine whether household composition has differential effects on the schooling of boys relative to girls, the interaction of each of the children composition variables with the gender of the child (*MALE*) dummy is included. There are several studies on child schooling where the effect of siblings in the same household has been examined. For example see Parish and Willis (1993), Morduch (2000) and Cornwell *et al.* (2005).

There are several adult characteristics, which affect the child's schooling. First, we include three dummy variables to indicate the mortality status of the parents, namely mother deceased, father deceased and both parents deceased. There is a fairly large

literature that examines whether child outcomes are different depending on whether the child lives with their parents or not. In Sub-Saharan Africa in particular, this is related to the widespread practise of child fostering (see for example Akresh, 2004). A child whose parents have died (either because of HIV/AIDS or otherwise) might be more likely to work, might experience psychological problems, or might suffer due to the disruption of having to live away from his or her schooling.

As pointed out earlier, there are several religions in Kenya. It has been argued that religion could have an effect on a child's schooling, so dummy variables for the main religions in Kenya are included. In addition, we included a dummy variable for household head literacy, taking a value of one if the house head has some primary school education and zero if not. This variable will capture the effect of the parent's or the household head's literacy on their children's schooling.

The other household variables used in the analysis include a set of province dummies. We did not include ethnicity variables since the provinces typically represent the demarcation of the ethnic groups in Kenya. Nairobi province is the base in the analysis. A dummy for the type of location of the household (urban or rural) was also included.

Violence in the family, between the mother and the partner/husband is quite likely to affect children.[28] We also include a variable for whether the child was hurt by either the parents or teachers.

[28] The violence variable included less than severe and severe violence where less severe was if the spouse ever pushed, shook or threw something to the wife; or if ever been slapped or twisted her arm; or punched with fist or something harmful; or ever been kicked or dragged. Severe was if the spouse ever tried to strangle or burn her; or ever been threatened or attacked with a knife/gun or other weapon.

Circumcision has been a cultural and religious activity carried out by almost all tribes in Kenya.[29] Cultures that practice child (both male and female) circumcision consider the process as a way of the child becoming an adult. Circumcised children are meant to feel and live as an adult whereas uncircumcised ones are considered to be still children regardless of their age. To a female child, it is considered to be a mark of beauty and a transition into womanhood and marriage and to the male child, is a transition to manhood. However, despite these cultural norms, there are disadvantages of the process. The rituals and celebrations take place during school times and this limits the children from attending their schools. In addition, there are potential medical consequences. For example, female circumcision has sexual consequences (Boyle 2002; Hoffman 2002; Obermeyer, 2003). The dummy variable for child circumcision was included to try and measure these effects.

In this study, we include the household wealth effect by including the wealth index or quintiles (poorest, poorer, middle, richer and richest). Wealth serves as a proxy for long-run household income or current household expenditure.[30] Filmer and Pritchett (2001) treat the asset index as a proxy for something not unobserved, a household's long-run economic status. This is in contrast to the earlier view and the study by Montgomery *et al*. (2000) where they used the asset index as a proxy for current expenditures. This variable was constructed using indicators of household assets by choosing appropriate weights which were determined by principal components analysis. Filmer and Pritchett

[29] Apart from Luo community, where the males are not circumcised.

[30] As noted by Filmer and Pritchett (2001), we do not use the indexed household assets as a parameter or measure for household current welfare or poverty since the conventional notion of poverty is based on the flow of consumption relative to some pre-determined poverty threshold, but by aggregating assets establishes only a measure of the household. In addition, the categorization used is based on the household's ranking within the distribution (a relative measure), whereas poverty thresholds typically are based on the expenditures essential for consumption of a determined bundle of goods.

(1999; 2001) established the validity of principal component analysis method in creating this index for wealth.

The alternative approach to capturing the effect of household wealth on child schooling is to simply enter all the asset variables separately in a linear multivariate regression equation. This procedure implicitly creates weights on the variables and handles the problem of controlling for wealth in estimating the impact of other, non-wealth, variables. The drawback of this approach is that the linear index of the assets using regression weights does not estimate the wealth effect because many assets exert both a direct and an indirect effect on outcomes. Therefore, although linear regression coefficients implicitly produce weights for the linear index of the asset variables that predicts the dependent variable most closely, there is no way to infer from these unconstrained coefficients the impact of an increase in wealth (Filmer & Pritchett, 2001).

Household wealth is an endogenous variable and there is no statistical software available to correct the bias caused by this endogeneity using an ordinal variable. To solve this problem, we re-constructed a continuous wealth variable using the principle component analysis (PCA) where the household assets were used. The first component was used as a proxy for the wealth variable. More about PCA is given in Chapter 2 section 2.4.3.

In this model we included an independent variable for any member in the household being diagnosed as HIV positive. This would give us an idea of the effect of this disease on children's school enrolment and attainment in Kenya and we might expect a negative effect of the disease on children's schooling. Here, we will need to include the sample selection model for the variable "Household member consent" defined by household,

given that an adult in the selected household had agreed to HIV/AIDS testing. Individuals who were observed were non randomly selected since they chose to be included into HIV testing or not, creating a sample selection problem which would lead to bias in our estimations. The solution to this is to estimate a sample selectivity model with the dependent variable being "Household member consent" to be tested for HIV. We included two instruments in this equation along with the exogenous independent variables from the main model. The first variable was *union* which is based on idea that if an individual had more than two unions, then she or he will be likely to refuse for HIV test because of fear to be found with HIV. The second variable was to capture if an individual *works away* from the household (family) mostly in the cities then he or she might have had an extra-marital relationship putting her or he in high risk of contracting HIV disease and so not willing to be tested. However, the opposite may be also true that the individual may want to know her or his HIV status probably to change the behaviour and so may agree to the test.[31]

As in chapter two, there are estimation issues in modelling child schooling. These issues include the household wealth endogeneity and the HIV sample selection. The issue of household wealth endogeneity is corrected by constructing a continuous wealth variable constructed via principal component analysis, enabling us to use the Rivers Voung procedure to correct for endogeneity of wealth (Rivers & Vuong, 1988). For HIV sample selection, we used two methods to correct the problem. The first one uses the Heckman procedure to deal with this non-random sample selection (Heckman, 1978; 1979). This is a standard selection procedure under the standard assumptions as outlined in section

[31] We need to note that data used in chapter two was based on individuals while in this chapter the data is based on household and therefore the instruments used for HIV are different.

3.5.3. More details of these two methods is given in chapter 2, section 2.4.2 – 2.4.4. The second method to solve the problem of the HIV missing data is the use of imputation by chain equations (ice). The issue of non-response in the data can be resolved using a variety of methods, including the available case method, weighting methods and model – based procedures such as maximum likelihood estimation. Overviews of such methods are given in Little and Rubin (1990), Schafer and Graham (2002) and Ibrahim *et al.* (2005). Some simple methods that are commonly used such as listwise or case deletion, pairwise deletion are found to be not adequate to compensate for nonresponse or bias (Allison, 2001; Schafer & Graham, 2002). For this analysis, we are concerned with one variable, the HIV variable with many missing data points meaning a sizeable loss of sample size.[32] To avoid this problem, we used the method of imputation by chain equation (ICE) to impute the missing data. Gabriele (2005) explains in detail the outline of the method and further evaluation criteria of imputation methods are described in Chambers (2003)[33].

The definitions and the list of all variables used in the models are given in Appendix 3.B.

3.5.3 Empirical Specification

In this section, we will give the specific empirical model used in the analysis of child schooling. The first equation is for the child's school attendance or enrolment which is a probit model given as:

[32] This was not the case in chapter two since it deals with the individuals who were HIV positive and not children in the household.
[33] See Rubin, 1976, 1987.

$$SchAtt_i = X_i^{'}\beta_1 + \gamma W_i + u_{1i} \quad (3.12a)$$

$$W_i = X_{Wi}^{'}\beta_2 + u_{2i} \quad (3.12b)$$

where X_i is the exogenous set of covariates comprising the individual (child) and household characteristics, β_j is the parameter vector we want to estimate, γ is a scalar representing the effect that W_i (variable measuring household wealth) has on $SchAtt_i$ for the child i and u_{ji} are the error terms.

Equation (3.12) is the base model we used to estimate the child's school attendance. In this estimation there are six model sets which compare the school attendance estimations based on different estimation issues. In each model set, we have given the estimations for child's school attendance (probit estimation), attainment (ordered probit) and grade progression rate (OLS).

The second model set deals with issues around the household wealth variable. Assume

$$W_i^* = X_{Wi}^{'}\beta_2 + u_{2i} \quad (3.13)$$

The wealth variable is given as a set of five binary indicators, allocating each household to one of 5 wealth categories from poorest to richest. In order to deal with the endogeneity of wealth, we used a bivariate ordered probit. The first equation is for the child schooling (attendance or attainment) and the second equation is an ordered probit for the ordinal wealth variable. In this case, equation (3.12 and 3.13) becomes a system of equations of latent variables where the error terms u_{1i} and u_{2i} are assumed to be jointly normal with correlation coefficient ρ, i.e. $\begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$

We observe the categorical variables for child's schooling (probit for attendance and ordered probit for attainment) and for wealth category as follows:

$$\left.\begin{array}{l} \underbrace{SchEnrol_i = \begin{cases} 0 \;\; if\; SchEnrol_i^* \le 0 \\ 1 \;\; if\; SchEnrol_i^* > 0 \end{cases}}_{School\;\; Attendance} \\ \underbrace{SchAtt_i = \begin{cases} 0 \;\; if\; SchAtt_i^* \le c_{11} \\ 1 \;\; if\; c_{11} < SchAtt_i^* \le c_{12} \\ 2 \;\; if\; c_{12} < SchAtt_i^* \end{cases}}_{School\;\; Grade\;\; Attainment} \end{array}\right\}, \quad \underbrace{W_i = \begin{cases} 1 \;\; if\; W_i^* \le c_{21} \\ 2 \;\; if\; c_{21} < W_i^* \le c_{22} \\ \vdots \\ 5 \;\; if\; c_{24} < W_i^* \end{cases}}_{Household\;\; wealth} \tag{3.14}$$

where c_{kl} and c_{km} are the cut-points to be estimated.

For grade attainment, the probability of observing $W_i = m$ and $SchAtt_i = l$ for individual i is given by

$$\begin{aligned} pr\left(SchAtt_i = l, W_i = m\right) &= pr\left(c_{2l-1} < SchAtt_i^* \le c_{2l}, c_{1m-1} < W_i^* \le c_{1m}\right) \\ &= pr\left(SchAtt_i^* \le c_{2l}, W_i^* \le c_{1m}\right) \\ &\quad - pr\left(SchAtt_i^* \le c_{2l}, W_i^* \le c_{1m-1}\right) \\ &\quad - pr\left(SchAtt_i^* \le c_{2l-1}, W_i^* \le c_{1m}\right) \\ &\quad + pr\left(SchAtt_i^* \le c_{2l-1}, W_i^* \le c_{1m-1}\right) \end{aligned} \tag{3.15}$$

If the distribution of u_{1i} and u_{2i} is bivariate standard normal with correlation ρ, then the individual contribution to the likelihood function is expressed as

$$\begin{aligned} pr\left(SchAtt_i = l, W_i = m\right) &= \Phi\left(\left(c_{2l} - \gamma_i X_{Wi}^{'}\beta_2 - X_i^{'}\beta_1\right)\xi, c_{1m} - X_{Wi}^{'}\beta_2, \rho\right) \\ &\quad - \Phi\left(\left(c_{2l} - \gamma_i X_{Wi}^{'}\beta_2 - X_i^{'}\beta_1\right)\xi, c_{1m-1} - X_{Wi}^{'}\beta_2, \rho\right) \\ &\quad - \Phi\left(\left(c_{2l-1} - \gamma_i X_{Wi}^{'}\beta_2 - X_i^{'}\beta_1\right)\xi, c_{1m} - X_{Wi}^{'}\beta_2, \rho\right) \\ &\quad + \Phi\left(\left(c_{2l-1} - \gamma_i X_{Wi}^{'}\beta_2 - X_i^{'}\beta_1\right)\xi, c_{1m-1} - X_{Wi}^{'}\beta_2, \rho\right) \end{aligned} \tag{3.16}$$

where $\Phi(.)$ is the bivariate standard normal cumulative distribution function with ξ and ρ defined as $\frac{1}{\sqrt{\gamma_i^2 + 2\gamma_i\rho + 1}}$ and $\xi(\gamma_i + \rho)$ respectively (see Appendix 3.11.1). This is a simultaneous bivariate ordered probit model. In the case where $\gamma = 0$, this model simplifies such that $\xi = 1$ and $\rho = \rho$ which becomes a seemingly unrelated specification.

The log-likelihood for an individual *i* is then given by:

$$\ln L_i = \sum_{l}^{L}\sum_{m}^{M} I\left(W_i = l, SchAtt_i = m\right) \ln Pr\left(W_i = l, SchAtt_i = m\right) \tag{3.17}$$

and with assumptions of independent observations, equation (3.17) is summed across all observations to obtain a log-likelihood for the entire sample of size N to get

$$\ln L = \sum_{i}^{N}\sum_{l}^{L}\sum_{m}^{M} I\left(W_i = l, SchAtt_i = m\right) \ln Pr\left(W_i = l, SchAtt_i = m\right) \tag{3.18}$$

Equation (3.18) is maximised using 20-point Gauss-Hermite quadrature in STATA 11.1. In the maximum likelihood estimation, ρ is not directly estimated, but *atanh* ρ is computed as

$$\text{atanh}\rho = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) \tag{3.19}$$

If $\rho = 0$, then the log likelihood for the bivariate probit models is equal to the sum of the log likelihoods of two univariate probit models. A likelihood-ratio test for $\rho = 0$ can be performed by comparing the likelihood of the full bivariate model with the

sum of the log likelihoods for the univariate probit models (Hardin, 1996, Van de Ven & Van Pragg, 1981).

In the discussion we have captured wealth effects on a child's schooling by including this set of dummy variables capturing the different wealth categories for the household where bivariate ordered probit model is used. The disadvantage with this approach is that it discards information about wealth – a continuous wealth index has been replaced with a set of 5 categories for comparison. There is a great deal of variation within these categories that is ignored here. For these reasons, we also include results with a continuous wealth variable where we have followed the DHS process, but stopped one step from the end. Namely, we use principal components analysis on the asset indicator variables to construct, using the first principal component, a continuous index of household wealth (see Chapter 2, section 2.4.3). Specifically,

$$\left.\begin{array}{l} W_i^A = pc\{X_{Ai}\}, \\ \text{the first principal component } \left(c_1 = X_{Ai}^{'}\alpha_{1i}\right) \text{ is used} \end{array}\right\} \tag{3.20}$$

c_1 is the first component extracted, α_{1i} is the regression coefficient (or the weight) for the observed household asset variable X_{Ai}, thus W_i^A is the variable measuring household wealth.

The endogeneity of wealth is dealt with using standard instrumental variables, with asset variables and other household characteristics (eg. Education level of household head) as instruments. The wealth equation becomes:

$$W_i^A = X_{Wi}^{'}\gamma + u_{3i} \tag{3.21}$$

Note in this case W_i^A is a continuous wealth variable given by a behavioural equation and we assume $Cov(u_{1i}, u_{3i}) \neq 0$, $E(u_{3i} \mid X_{Wi}) = 0$. To correct for bias due to the wealth endogeneity, we use standard Rivers-Voung procedure since W_i^* is now continuous. This involves obtaining the residuals from the ordinary least squares regression model of the continuous wealth variable on the X_{wi} regressors, then using these residuals as an additional regressor in the main equation.

The other problem we face in estimation is with the variable capturing whether an adult in the household is HIV positive, which creates a problem of sample selection. The data could be assumed to be missing at random or missing completely at random. Due to this uncertainty, two methods are used to correct the bias due to this HIV sample selection[34]. Firstly, we used a version of the Heckman procedure where we specify:

$$SchAtt_i = X_i^{'}\beta_1 + W_i^{A'}\beta_2 + u_{1i} \quad (3.22)$$

$$T_i^* = X_{Ti}^{'}\psi + u_{4i}, \text{ and a person's HIV status is observed if } T_i^* > 0, \quad (3.23)$$

T_i^* is the latent variable capturing propensity to be tested for HIV. The potential bias in selection is represented by the assumption $Cov(u_{4i}, u_{1i}) = \sigma_{u_4 u_1} \neq 0$. Since the DHS survey collects information about the individuals who choose not to be tested for HIV then it becomes possible to estimate equation (3.23). The covariates in equation (3.23), X_{Hi}, include factors which would influence the individual's choice regarding HIV testing. For identification, at least one variable in X_{Hi} needs to not belong in equation (3.22) - factor(s) that influence individuals choice to be tested, but not the

[34] See previous section 3.5.2.4 for more explanation.

child's schooling equation except through the impact on the individual choice to be tested or not.

Following the standard Heckman procedure (see Wooldridge, 2006, p. 618-620), equation (3.12a) can be re-written as

$$SchAtt_i = X_i^{'}\beta_1 + W_i^{A'}\beta_2 + \beta_3\lambda\left(X_{Ti}^{'}\psi\right) + \varepsilon_i \tag{3.24}$$

An estimation of the parameters of equation (3.23) can be used to estimate the inverse Mills ratio, $\lambda\left(X_{Ti}^{'}\psi\right)$, and this would then be included as an extra variable in the estimation of equation (3.22).

The second method to solve the problem of the HIV missing data is the use of imputation by chain equations. One variable, HIV status, has many missing data points meaning a sizeable loss of sample size in our estimation of (3.12). To avoid this problem, we used the method of imputation by chain equation (ICE) to impute the missing data. Assuming the data are *missing completely at random* (MCAR), we specify:

$$H_i = \begin{cases} 1\ if\ \ X_{Hi}^{'}\psi > 0 \\ 0\ if\ \ X_{Hi}^{'}\psi < 0 \end{cases} \tag{3.25a}$$

and so

$$H_i = \begin{cases} H_i \text{ if } H_i \text{ is observed} \\ H_i\ otherwise \end{cases} \tag{3.25b}$$

where H_i is our observed variable. Equation (3.25) describes the equation by which the relationship between HIV status and observable variables is estimated for those adults in the household. X_{Hi} is observed for all (or at least most) households, and also includes some variables not in equation (3.22). Based on estimates from (3.25), the

HIV propensity is estimated for other adults who were not tested, and an imputed value of HIV status obtained. For our analysis we used STATA software where we use the procedures written by Royston (2004, 2005).

This set of equations could in principle be estimated as a system using maximum likelihood, but given the complexity of the system, with at least four equations to be estimated and nonlinearities present in different forms, it is likely that we would experience severe convergence problems. Integration of a likelihood function across four dimensions is computationally intensive, and often unreliable (Arendt & Holm, 2006). Consequently we have opted for the slightly more ad hoc, but hopefully more reliable, approach of augmenting the main equation with the appropriate residuals to give consistent estimates. While there is a small loss in efficiency compared with maximum likelihood, the estimation is much less vulnerable to convergence and stability problems.

3.6 Results

3.6.1 Descriptive Statistics for Selected Variables

In our data analysis, we used STATA/IC 11 for Windows, (2010) to obtain the model estimates. Table 3.2 gives descriptive statistics for some selected variables. From this table, there is a large difference between attendance rates for primary and secondary school aged children, 86% and 61% respectively. Of 3340 children aged 15 – 18 years old, children with some but not all primary schooling have the highest proportion at about 56%, followed by those who have never enrolled in school with 29% leaving about 15% to those who have completed primary school. About 68% of children successively progressed from one grade to the next in a given year during their schooling period. The total number of school-aged children in the sample was 9515, with 51% being male.

Looking to other household characteristics, for 12% of the children (aged 6 – 14 years old) their father is not alive, 5.6% have a deceased mother and about 3% have both parents deceased. The proportion of children with household head having some primary school education is 90% for secondary school aged children compared to 72% for primary school children The proportion of female children being circumcised is 41% among the primary school aged female children and 33% for secondary school-aged children.

The various models used are given in Table 3.1 below.

Table 3.1. **Summary of Equations used in the Models.**

Model set	Equation used (refer to section 3.5.3)	Explanation
I	3.12	• Does not deal with endogeneity of household wealth: Assumes $\rho = 0$ • No HIV variable include
II	3.12a, 3.12b and 3.14	• Use bivariate ordered probit to correct for wealth endogeneity • Use predicted probabilities as proxies for wealth in school grade progression • No HIV variable included
III	3.20, and 3.21	• Uses Instrumental variables to correct for wealth endogeneity • No HIV variable included
IV	3.22, 3.20 and 3.21	• Same as Model set III, but • Uses only observed data for any household member HIV/AIDS status
V	(3.22, 3.23, 3.20 and 3.21) = 3.24	• Same as Model set IV, but • A Heckman selectivity Model was fitted for any household member agreed for HIV testing (both tested and not tested) • IMR computed for school grade progression • IV regression used for school grade progression
VI	3.24, 3.25b, 3.20 and 3.21	• Uses Instrumental variables to correct for wealth endogeneity • Use observed and imputed data for household members with HIV/AIDS. • IV regression used for school grade progression

3.6.2 Current School Enrolment

The estimation results for school enrolment are based on six different model sets.

3.6.2.1 Model set I: Assuming Wealth Exogeneity and no sample Selection Bias

The results for current school enrolment are given in Table 3.3 with the dependent variable representing whether the child is currently enrolled in school or not. Here we have estimated models separately for primary school aged children (6 – 14 years old) and secondary school aged children (15 – 18 years old). The table gives the probit coefficients with robust standard errors and marginal effects. The marginal effects are defined as partial derivatives of the probability of a child being enrolled in school with respect to the individual control variables, holding all dummy variables to zero and all other variables at the sample means.

From these estimates, several factors have a significant effect on school enrolment. The dummy variable for gender reveals no significant difference between school enrolment of primary school aged male and female children. However, in the model for secondary school aged children (*column 4 and 5*) we find a significant coefficient of with marginal effect of 0.357, which implies the male child in the household is much more likely to enrol in secondary school than a female.

Looking into the effect of presence of other children in the household, the results shows little influence on primary school enrolment, apart from the presence of at least one primary and secondary school aged male in the household. The variable for primary

school aged child has a coefficient of -0.064 ($p<0.1$) and marginal effect of -0.004. This means a female child is about 0.4 percentage points less likely to be enrolment in secondary school if there is at least one male of the same age in the household. This indicates a very small element of rivalry among the primary school aged children in the same household. In contrast, the presence of at least one secondary aged male in the same household has a positive significant coefficient of 0.135 ($p<0.01$) and marginal effect of 0.009. This means a female child is about 1% more likely to be enrolled in primary school if there is at least one secondary school aged male child in the same household. This indicates some synergy effect between younger female and older male children in the same household. Note that the interaction variables of these children with a male child have insignificant coefficients.

There is more sibling effect in the case of secondary school enrolment than in the primary school model. Here, we find the variables for both male and female children aged 0 – 5 years old have negative coefficients of -0.200 and -0.153 ($p<0.01$) with marginal effects of -0.077 and -0.059 respectively. These mean a female child is about 6-8 percentage points less likely to be enrolled in secondary school if there is at least one pre-primary school aged child in the same household. This is expected because the older female children stay home taking care of the younger siblings. However, the interaction variable for pre-primary school aged male with an older male cancels out the marginal effect. The interaction variable for the pre-primary school aged females with an older male child nets out the marginal effect to 0.04. This implies the presence of at least one pre-school female in the household reduces the chance of an older male being enrolled in secondary school by about 4 percentage points, indicating some effect, but much weaker than that for girls.

The variable for primary school aged male child has a coefficient of 0.162 ($p<0.01$) and marginal effect of 0.063. This implies a female child is about 6 percentage points more likely to be enrolled in school if there is at least one primary school aged male in the same household. This shows some synergy effect for female and male children in the same household in favour of the older females. However, the marginal effect of -0.055 for the interaction variable of the primary school aged male with an older male cancels out the effect. The variable for a primary school aged female has a coefficient of 0.159 ($p<0.01$) with a marginal effect of 0.062. This means a female child will be about 6 percentage points more likely to attend secondary school if there is at least one primary school aged female in the same household. Again this shows some synergy effect among children in the same household in favour of the older females. The reason for this could be that a primary school aged female child helps with household chores enabling an older female child attend school.

The variable for a secondary school aged female child gives a very strong effect with a coefficient of 0.281 ($p<0.01$) and marginal effect of 0.108. This means a female child is about 11 percentage points more likely to attend secondary school if there is at least one female child of the same age in the same household, a very strong synergy effect. However, the interaction variable with a male child has a marginal effect of -0.203 netting out this effect to -0.095 for a male child. This means that a male child will be about 10 percentage points less likely to attend secondary school if there is a female child of the same age in the same household. This indicates some rivalry between the secondary school aged males and the females in the same household, where male children are disadvantaged in school attendance. Recall the initial results that the male dummy was positive and significant implying boys are more likely to go to school, but

this effect suggests this difference is smaller when there is a secondary aged girl in the household.

Turning to adult characteristics, we find the variable for both parents being deceased has a coefficient of -0.678 ($p<0.01$) with a marginal effect of -0.081 for the primary school model.[35] The dummy variable for household religion has a coefficient of 0.387 and marginal effect of 0.027 for Protestant and Seventh Day Adventist (SDA). This means a child from the Protestant and SDA household is about 3 percentage points more likely to be enrolled in primary school compared to a child from a catholic and other or no religion households.

Turning to the dummy variables for household wealth, for the primary school attendance model we find all four dummies with significant coefficients and marginal effects of 0.013, 0.022, 0.022 and 0.026 for poorer, middle, richer and richest household respectively. These shows that as the household get wealthier a child will be more likely to attend primary school although marginal effects are not very large. In the case of secondary school attendance, we get a surprising result with a negative coefficient of -0.356 ($p<0.01$) and marginal effect of -0.140 for the richest household. This means a child from the wealthiest household category is about 14 percentage points less likely to attend secondary school compared to a child from the poorest household category. This effect is seen in Table 3.6 where compared to the poorest household, the marginal effect increases in magnitude from poorer to the richest household. The reason for this effect could not be explained by this study. However, a possible reason might be the fact that at this age the children are in their teenage years and being from the rich households,

[35] There were no observations for the secondary school aged children.

they are more likely to engage in other activities and not studies. These might be use of drugs, alcohol, and even pregnancies, etc. which, leads them to drop out of the school.

The variable for the household head having at least some primary education gives significant results with marginal effects of 0.443 and 0.702 for primary and secondary school attendance respectively. This means a child from a household whose head has some primary education is about 44 and 70 percentage points more likely to attend primary and secondary school respectively, compared to a child whose household head is not educated.

The dummy variable for urban area has a negative coefficient of -0.207 ($p<0.1$) with marginal effect of -0.081 for secondary school attendance. This means an older child living in the urban area is about 8 percentage points less likely to attend secondary school. The dummy variables for the provinces have some effect in the primary school attendance model. The Central, Coast, Eastern, Nyanza and Western provinces have significant coefficients of 0.384, 0.314, 0.504, 0.802 and 0.581 and marginal effects of 0.020, 0.017, 0.024, 0.032 and 0.026 respectively. These means a child residing in any of the provinces is about 2 -3 percentage points more likely to attend primary school compared to a child residing in the Nairobi province.

In the secondary school model we find the variable for the Western province has a significant marginal effect of 0.131. This implies a child from this province will be about 13 percentage points more likely to attend her or his school compared to a child residing in the Nairobi province.

Another variable included in the model is an indicator for whether the household had incidents of violence, with a significant coefficient of -0.163 and marginal effect of -0.064 for the secondary school children. This means a child from a household with violence is about 6 percentage points less likely to attend secondary school compared to a child from a household with no reported violence. The variable for a polygamous household has a coefficient of -0.291 ($p<0.01$) and marginal effect of -0.024 for the primary school attendance model. This implies a child from a household with more than one wife is about 2 percentage points less likely to attend primary school compared to a child from a monogamous household. Lastly we find a weakly significant effect female circumcision with a coefficient of -0.131 ($p<0.1$) and marginal effect of -0.009.

3.6.2.2 Model Set II: Allowing for Endogenous Wealth Variable (ordinal)

The next set of results uses IV estimation to correct for the endogeneity of wealth. Results for current school enrolment are given in Table 3.6. The models are for primary school attendance using 8,270 children aged 6 – 14 years (*column a)* and secondary school attendance using 2,894 children aged 15 - 18 years (*column b*).

The models give similar results to that in the Model set I apart from few variables with different coefficients and marginal effects magnitudes. The dummy variables for household wealth have significant negative coefficients for secondary school attendance, whereas these were not significant in Model set I.

The Wald test for exogeneity of wealth shows a significant ρ with χ^2 = 3.01 (p = 0.083) and χ^2 = 7.18 (p = 0.007) for primary and secondary school attendance

respectively. This means rejecting the hypothesis of wealth exogeneity, implying estimating the models with OLS will give unreliable estimates.

The equation for wealth is shown in Table 3.7. The instruments in the models include occupation dummies, with the dummy variables for professional, technical or manager having coefficients of 0.758 and 0.824 ($p<0.01$) for primary and secondary school enrolment respectively. This means these occupations in this category are more likely to make a household wealthier. A similar effect is found in the occupation variables for clerical, sales, household and domestic, and services. The variables for skilled manual occupation has a significant coefficient of 0.444 for secondary school and unskilled manual work has a coefficient of 0.164 ($p<0.01$) for the primary school model. The dummy variable for self-employed in agriculture has a negative coefficient of -0.226 ($p<0.01$) for secondary school enrolment. This means self-employment in agriculture is associated with lower household wealth.

The other instruments in the wealth equation are the dummies for ethnic tribes. The Embu tribe have significant coefficients of 0.842 and 0.838 for the primary and secondary school models respectively. This means an Embu household is more likely to be wealthier compared to other tribes. A similar effect is found in the variables for the Luo and Meru tribes with significant coefficients of 0.526 and 0.710 for primary school respectively and coefficients of 1.015 and 0.637 for secondary school enrolment. Other significant effects are found for various tribes, suggesting quite a degree of variation in wealth across tribes.

The variable for the household head's education has coefficients of 0.361 and 0.371 for primary and secondary school children respectively. A household headed by an

individual with at least some primary school education is likely to be wealthier compared to a household headed by uneducated person. The dummy variable for urban area has coefficients of 1.508 and 1.521 ($p<0.01$) for the primary and secondary school models.

3.6.2.3 Model set III: IV Probit Model for Endogenous Wealth Variable (Continuous)

In the next set of results, wealth is included via a continuous index rather than the set of dummies for wealth categories. Instrumental Variable Probit is used to account for the endogeneity of wealth. This continuous wealth variable was constructed using principal component analysis (PCA) from household assets. From the PCA, the first component was retained because it had an eigenvalue of 4.0 meaning it accounted for variance far greater than 1, which means it accounts for more variance than any of the original observed variables. To confirm the legitimacy of using just the first principal component, we plotted a scree plot. Normally the components before the point where the elbow or sharp curve occurs are retained. The elbow is found at the second component suggesting that first component accounted for a large amount of the combined variance (see Figure 3.1). Bartlett test is used to test for intercorrelation of the variables used and it reveals rejection of the null hypothesis ($\chi^2 = 1.1*10^5$, $p-value = 0.000$). The Kaiser-Meyer-Olkin (KMO) statistic is used to measure the sampling adequacy at predicting the data if the data are to factor well based on correlation and partial correlation. The variables used in the PCA had KMO value of 0.882.

This model is structured the same as in Model set I. Table 3.11 gives the ivprobit coefficients with robust standard errors and marginal effects. *Column a* show the results for the primary school using 8,209 children and *column b* for the secondary school using 2,867 children.

The results are similar to that in Model set I except for a few variables. In this model there is some evidence that primary children with a deceased father are more likely to attend school. This is not a clear effect, though, and is based on a relatively small number of cases. The fact that it is not robust across models also suggests it is not a strong effect.

The dummy variable for urban area has negative coefficient of -0.425 ($p<0.01$) and marginal effect of -0.052 for the primary school model. This means a child is about 5 percentage points less likely to attend school if they come from an urban area, compared to a child from rural areas.

The Wald test for the endogeneity gives significant results with $\chi_1^2 = 15.890$ and $p = 0.000$ for primary school attendance but insignificant for secondary school attendance. This means that ignoring endogeneity in the case of the primary school model will give unreliable estimates.

3.6.2.4 Model set IV: Including HIV/AIDS as an Explanatory Variable

In this section of results we have included a variable for whether any adult member in the household has HIV/AIDS. The results are given in Table 3.14, *column a* showing the

results for 3,604 children aged 6 - 14 years and *column b* the results for 1,219 children aged 15 - 18 years.

The model has similar results to Model set I with only a few differences. One of the differences is in the sibling variables. We find significant effects of pre-primary aged children. However, in this model, the effects are not strongly significant (p-values are not less than 1 %), so we do not place too much stock on this result, and retain the conclusion found earlier that the presence of pre-primary aged children can have a detrimental effect on girls at secondary age, but not at primary school age.

The variable for HIV/AIDS give a very significant and negative coefficient of -0.417 and marginal effect of -0.162 for the secondary school model. This implies a child is about 16% less likely to attend secondary school if there an individual in the household who is HIV positive. This is a strong effect. It is quite plausible, as the older child in such household is going to take more of the household responsibilities, including taking care of the sick person. In addition, there could be some diversion of household income to patient care and treatment, away from paying the school expenses for the child.

The tests for endogeneity of household wealth gives same results as in Model set III.

3.6.2.5 Model set V: Sample Selection Model for Missing Data on HIV Status

This model uses the same structure as in Model set IV except that we seek to deal with the sample selection issues associated with a high proportion of the sample not being tested for HIV status. To correct for possible bias due to sample selection on HIV

testing, we used Heckman's procedure. The results are given in Table 3.17. *Column a* shows the results for 4,009 children aged between 6 and 14 and *column b* for 1,368 children aged between 15 and 18 years. The results for the selectivity model are given in Table 3.18.

Although the results in this model set are similar to the previous Model sets I and IV, there are notable differences in a few variables in significance and magnitudes of the coefficients and marginal effects. Compared to the Model set I, we find the interaction variable for pre-primary school aged males with a male child has an insignificant coefficient. The variable for primary school aged male has a significant coefficient of -0.152 and marginal effect of -0.008 for primary school. This means a female child is about 1% less likely to attend primary school if there is at least a primary school aged male child in the same household. We find the interaction variable for pre-primary school aged female children with a male child has a significant coefficient of 0.235 and marginal effect of 0.012 for the primary school model. This means a male child is about 1% more likely to attend primary school if there is at least a primary school aged female in the same household. There are other small differences in sibling effects, but no substantial pattern of difference.

Unlike in the Model set I, the variable for both parents deceased now has an insignificant coefficient.

The HIV variable in this model has similar results to that in the Model set IV with a negative coefficient of -0.412 ($p<0.01$) and a marginal effect of -0.159 for the case of child's secondary school attendance. This implies a secondary-aged child is about 16

percentage points less likely to attend school if any of their household members is positive.

Overall, the allowance for sample selection has made very little difference to the significance and magnitude of effects.

3.6.6.6 Sample Selection Model

The selection model considers the decision of household members who were selected for HIV testing whether to agree to be tested. There are likely to be unobservable factors that influence individuals to make this decision. The Heckman procedure (Heckman, 1978; 1979) includes an inverse Mills ratio (IMR) to control for the unobservable factors that contribute to the selection decision and are correlated with schooling decisions. The likelihood Ratio test is used to test for the presence of such unobservables that are correlated between the selection and the main equation. If the correlation ρ is zero then the selection equation is not required. The result shows that for primary school model, $\rho = 0.262$ and is marginally significant with $\chi_1^2 = 3.09$ $(p = 0.079)$ and for the secondary school model $\rho = 0.626$ with $\chi_1^2 = 3.56$ $(p = 0.059)$. These results confirm that by ignoring the sample selection issue the estimates could lead to some bias.

In the selection model, the instruments include a dummy variable for the number of unions an individual had and a dummy for if the individual was working away from the family. The variable for union is found to have significant coefficient of -0.517 ($p<0.1$) for the secondary school selection model. The variable for working away from the family had coefficient values of -0.302 and -.0396 ($p<0.01$) for primary and secondary

school models respectively. These indicate an individual who works away from the family is less likely to agree to be tested for HIV status.

There is some weak evidence of household composition effects that appear to influence the individual's decision, especially for the secondary school model. For example, an individual from a household with pre-primary school aged female is more likely to go for HIV testing.

The variables for religion indicate that those from Protestant and Seventh Day Adventist religions are more likely to go for HIV testing. The variable for an individual residing in the urban areas suggests that an individual from the urban areas is less likely to accept HIV testing compared to one living in the rural areas.

Lastly, the variable indicating violence in the household has a coefficient of 0.375 $p<0.01$) in the model with primary school aged children, indicating an individual from a household with violence is more likely to go for HIV testing.

3.6.2.7 Model set VI: Use of Imputation by Chained Equations for HIV/AIDS Missing Data

In this Model Set VI we use a different approach from Model set V to account for the sample selection problem. In this model set impute the missing observations in the HIV/AIDS data using imputation by chain equations (ice) (see section 3.5.3 for more details).

Generally, the estimates in this model are very similar to that in Model set V, the difference mostly is in the magnitudes and significance of a few variables. Note that in

this model set the standard errors are smaller due to a larger number of observations, which has some impact on the significance of the estimates.

The results are given in Table 3.21. The table gives the ivprobit coefficients and the respective marginal effects with robust standard errors. *Column a* use 8,209 children aged between 6 and 14 years old and *column b* use 2,867 children aged between 15 and 18 years. The large number of observations compared to the Model set V is from the imputation of the missing observations in HIV/AIDS variable.

Generally, compared to Model set V we see similar effects except a few worth mentioning. Firstly, due to larger number of observations, standard errors are smaller resulting to different significance of estimates. In this model, the interaction variable for primary school aged female children with a male child has an insignificant coefficient. The variable for secondary school aged female children suggest a stronger sibling synergy effect whereby a female child is about 11% more likely to attend secondary school if there is another a female child of the same age category in the household. However, this effect is largely netted for a male child by the interaction variable with a male child. In fact a male child is about 9% less likely to attend secondary school if there is a female of the same age in the household. This indicates a rivalry effect among these children with a male child being disadvantaged in secondary school attendance.

The variable for both parents deceased has a negative significant coefficient of -0.687 ($p<0.01$) with marginal effect of -0.113 for the primary school model, a much stronger effect than in model set V.

The variable capturing the effect of a household member being HIV positive has the same effect as in model set V with a significant coefficient of -0.205 and marginal effect of -0.081 for secondary schooling. This is quite a substantial effect.

3.6.3 Grade Attainment

3.6.3.1 Model set I: Assuming Wealth Exogeneity and no sample Selection Bias

The basic model for grade attainment is estimated using the ordered probit regression and the results are given in Table 3.4. The sample covers children aged 15- to 18 years old. The dependent variable is grade attainment for the children aged 15 - 18 years old. A positive coefficient on a particular explanatory variable implies that the variable increases the probability of the child having completed primary school, implying a lower probability of having not attained any schooling. The table gives coefficients and marginal effects (standard errors in parentheses) computed at sample means for continuous variables and zero for dummy variables.

First, age has no significant effect on the probability of primary school completion, implying that after a child reaches 15 years of age, and has not completed primary school, it is likely they will never finish. The dummy variable for gender gives a very statistically significant coefficient, and marginal effects suggest that a male child is 17 percentage points more likely than a girl to have not gone to school, about 9 percentage points less likely to have completed some primary school and about 8 percentage points less likely to complete primary school.

Looking to the effect of other siblings in the same household, we find the variable for primary school aged male child is significant with a coefficient of -0.074 ($p<0.01$). The interaction term with gender of the child is, however, significant with a similar but opposite sign. These tell us that a girl with a male sibling who is currently at primary age is much more likely to not have completed primary school. This is mostly because when the girl was at primary school age, she was caring for her younger sibling, who at that point would have been of pre-primary age. Older boys do not seem to suffer the same detrimental effect on their schooling.

There is weak evidence that the presence of pre-primary female child has a positive effect on school attainment, although the effect is not there for pre-primary aged boys. The other major sibling effect is that the variable for secondary school aged female child has a coefficient of -0.187 ($p<0.01$) with marginal effects of 0.067, -0.037 and -0.030 for no schooling, with some primary schooling and completed primary school respectively. This means a female child is 3 percentage points less likely to have completed if there is at least one other secondary school aged female child in the same household. Again, this indicates some rivalry among the female children of the same age group in their primary schooling. However, these marginal effects are largely netted out for a male child, suggesting this rivalry affects girls but not boys.

The variable for the household head's education has very strong positive effect. The associated marginal effect of 0.161 for completing primary school means a child is 16 percentage points more likely to complete her or his primary school if the household head has some primary school education.

The dummy variables for urban area and provinces are found to be insignificant apart from the Nyanza and North Eastern provinces with significant coefficients of -0.342 and -0.317 respectively. The associated marginal effects of -0.047 and -0.043 means a child from in these provinces is about 5 and 4 percentage points less likely to have completed primary school compared to a child from Nairobi province.

The last variable with a significant effect is for household violence, with a coefficient of 0.121 ($p<0.1$) and marginal effect of 0.021 for completing primary school. This suggests a child from a household with some violence has about 2 percent points more likely to have completed primary school, although the effect is only barely significant

Note there is no evidence of a wealth effect at all on the ability of these children to complete primary school.

3.6.3.2 Model Set II: Allowing for Endogenous Wealth Variable (ordinal)

The model is same as that in Model set I above but using the bivariate ordered probit regression model to deal with the endogeneity of wealth. The results are given in Table 3.8.

The model shows very similar results as in the Model set I except for small changes in the magnitude of coefficients and marginal effects. The only notable difference is found in the variable for household violence, which now does not show a significant effect. The test for correlation of errors across equations indicates no evidence for endogeneity of wealth, with $\chi^2 = 0.97$ ($p>0.1$).

3.6.3.3 Model set III: IV Probit Model for Endogenous Wealth Variable (Continuous)

The structure of this model is same as that in the Model set II except we now include wealth as a continuous variable, use an IV probit to correct for the endogeneity of wealth. The grade attainment estimates given by the ordered probit regression are shown in Table 3.12.

This model gives similar results to Model set I with a few exception. First, we now find the variable for household wealth has a significant coefficient of -0.119 and marginal effects of 0.043, -0.024 and -0.019. This suggested that for every one unit increase of household wealth a child is about 2 percentage points less likely to have completed primary school. This is not what we expect because as wealth increases, we expect better resources for the child's schooling so the effect could be due to other factors correlated to wealth.

Another variable with different effect is for Central province which is now significant with a coefficient of -0.345 ($p<0.05$). The marginal effect suggests a child is about 5 percentage points less likely to have completed primary school if they are from Central province compared to a child residing in Nairobi province. The same magnitude of effect is found in the provinces of Western and North Eastern.

3.6.3.4 Model set IV: Including HIV/AIDS as an Explanatory Variable

The model has the same structure as in Model set III but with the addition of a variable indicating if there is an adult in the household who is HIV positive. The results are given in Table 3.15.

The model has mostly very similar results to those in Model set I; here we highlight a few different results. Firstly, the dummy variable for gender now has an insignificant coefficient. Also, the variables for primary school aged males and for interaction with a male child have insignificant coefficients. However, we find the variable for pre-primary school aged female children has a significant coefficient of 0.156, where previously it was only marginally significant. Marginal effects suggest a female child has about 3 percentage points better chance of having completed primary school if there is a at least one pre-primary school aged female child in the same household. This effect is cancelled out by the interaction variable with a male child. The variable for the presence of primary school aged females has a coefficient of -0.096 ($p<0.1$), which shows a small negative effect as in some earlier model sets.

The variable for household wealth and the dummy variables for the Nyanza, North Eastern provinces, and household violence have insignificant coefficients, as does the newly introduced variable for HIV/AIDS.

3.6.3.5 Model set V: Sample Selection Model for Missing Data on HIV Status

The structure of this model is same as that in the Model set IV, except we used the Heckman procedure to correct for possible sample selection bias due to the presence of many missing values on the HIV variable. To do this an Inverse Mills ratio was constructed and included in the equation. The results are given in Table 3.19.

The Inverse Mills ratio has a significant coefficient of 2.079 ($p<0.1$). This indicates the presence of unobservables driving selection that also affect educational attainment, and ignoring the sample selection would give biased estimates.

Compared to Model set IV, where the sample selection bias is ignored, virtually all results are very similar. While the non-random selection is clearly significant, it does not have a big impact on the parameter estimates.

The main exception to this is for the variable indicating violence in the household which now has a significant coefficient of 0.351 ($p<0.01$) and marginal effect of -0.11, 0.040 and 0.071. In Model set IV this variable is not significant. The result implies that a child from a household with violence is about 7 percentage points more likely to have completed primary school. This is a surprising result.

The variable for HIV/AIDS continues to prove insignificant, as does the measure of household wealth.

3.6.3.1 Model set VI: Use of Imputation by Chained Equations for HIV/AIDS Missing Data

The final set of results for grade attainment uses imputed values for the HIV/AIDS variable when the data is missing. This allows us to us get a much bigger sample size. Ultimately this is the preferred model, as it deals with the endogeneity of wealth as well as missing data. The results are given in Table 3.22, with a sample of 2,867 children.

These results are more similar to those in Model set I and differ more from Model V, probably due to the sizeable improvement in the number of observations used in estimation. As with Model set I, we find males have a much lower chance of completing some or all primary school. The presence of primary or secondary aged children reduces the chances of girls completing primary education, but not boys. As with all models, a more educated household head increases substantially the chances of a child completing primary school. The other result that seems to apply fairly consistently across models is the lack of effect of HIV/AIDS in the household on children's school attainment. We note that in this model set, wealth actually has a strong negative effect, a finding that is unique to this model set. This is in contrast with results on attendance, where wealth has a strong positive effect on attendance. This result here suggests that better attendance associated with wealth does not necessarily translate into a higher chance of wealthier children completing primary school.

3.6.4 Grade Progression

3.6.4.1 Model set I: Assuming Wealth Exogeneity and no sample Selection Bias

In this section, we obtain estimates for model of the rates of grade progression for children aged 6 – 18 years old who were currently enrolled in primary and secondary school. We have separately analysed models for overall, rural, urban, female and male. Results for the base case where wealth is treated as exogenous and the HIV variable excluded are given in Table 3.5, showing the coefficients and associated robust standard errors (in parenthesis).

Starting with the age variable, this has a negative coefficient of around -0.12 and for the overall model and the various sub-models. It is highly significant. The variable for squared age is also very significant. This suggests a U shaped relationship between age of the child and the child's rates of grade progression. As the age of the child increases, the grade progression falls to a minimum value at the age of about 13 years of age after that the rate of grade progression increases with age. The decline through the early years is indicative of students falling further and further behind the ideal progression of one grade per year. Those who are still in school beyond the age of 13 show better overall progression, reflecting that this is probably the stronger, self-selected group of students.

The dummy variable for gender has negative coefficients of -0.032 and -0.048 ($p<0.01$) for overall and rural models respectively. This means male rates of grade progression in the whole country and in the rural areas are significantly slower that that of a female

child. Interestingly, this is not the case for the urban sample, where there is no significant difference between male and female progression rates.

The other variables included in the models with significant effects include the presence of siblings in the same household. The variable for the pre-primary school aged male child has significant coefficients of -0.07, -0.028, -0.039, -0.030 and -0.031 ($p<0.01$) for overall, rural, urban, female and male models respectively. These means the presence of at least one pre-primary school aged male child in the household will cause a child's rates of progression to be significantly slower. The interaction term for gender suggests this is the case for girls, but that boys' progression rates are unaffected by the presence of pre-primary aged boys. The presence of pre-primary school aged girls in the same household decreases the rates of grade progression for an older female in the country by 0.012 and in the rural by 0.018, not being significant in the urban areas. The variable for primary school aged males has a coefficient of -0.007 ($p<0.1$) for the overall sample, -0.024 ($p<0.05$) for the urban areas and -0.008 ($p<0.1$) for a female child model. This suggests lower rates of grade progression by a small amount over the whole sample and by 0.024 in the urban areas if there is a primary school aged male child in the same household. Again, the interaction variable with a male child nets out these negative effects. It appears the strongest negative effect is on girls caring for pre-primary aged children, with a weaker effect for primary aged children.

The variable indicating if the child's father is deceased has significant coefficients of -0.023 and -0.039 for the overall and rural models respectively, meaning a child's progression rates will be significantly lower. It appears households are less able to cope with this loss in the rural areas than the urban areas. Notably, the loss of the mother does

not have any significant effect, suggesting it is the loss of access to resources (income) through loss of the father that has a greater impact.

Turning to differences between religions, the only significant effect is for Islam. Estimates suggest a Muslim child will progress through the grades more slowly in the entire sample, with a very strong negative effect of 0.098 in the urban areas.

The dummy variables for household wealth give very significant coefficients in all models except the urban model. For the overall model, the coefficients are 0.047, 0.061, 0.126 and 0.179 for the poorer, middle, richer, and richest households. This means a child from the richest household will have grade progression of 0.179 grades per year higher than the poorest, followed by the a child from richer with a rate of 0.126, then that from middle-wealth household with a rate of 0.061 and least a child from poorer household with a rate of 0.047. A similar effect is found in the rural, female and male models. In the urban model, we find slightly different results. The richest household has a very significant coefficient of 0.134 and middle household with a coefficient of 0.071 ($p<0.1$), but not significant differences for the second and third wealth categories. Essentially, the wealth effects on progression are strongest in rural areas, and still present but weaker in the urban areas.

The variable for household head education has very significant coefficients for all models. For example in the urban the model has a value of 0.217 ($p<0.01$). This effect is similar across all models – the whole sample, females and males, urban and rural, indicating a sizeable benefit if the household head has some primary school education.

Turning to geographical factors, the dummy variable for urban residence has significant coefficients of 0.022 and 0.032 for the whole sample and male models respectively. This implies a child from an urban area will have a higher rate of grade progression, especially boys. The dummy variable for provinces suggest there are several with significantly poorer progression rates than the base (Nairobi), namely Coast, Estern and Western province, while Central province appears to do better.

The variable for household violence has coefficients of -0.026, -0.024, -0.038 and -0.037 ($p<0.01$) for the overall, rural, urban and female models respectively. This means that violence in the house has a strong negative effect on the rate of grade progression, with the strongest effects being in urban areas and on females.

3.6.4.2 Model Set II: Allowing for Endogenous Wealth Variable (ordinal)

Table 3.10, shows results for the grade progression model allowing for the endogeneity of wealth where predicted probabilities of being in each wealth category (based on a separate ordered probit of wealth) are used.

In this model we find the household wealth dummy variables have a very strong effect on rates of grade progression. When wealth dummies were included, wealth shows as having very strong positive effect on grade progression, but in this model, virtually none of the wealth effects are significant. This is surprising, and we suspect has more to do with how the endogeneity is dealt with. We will see in later sections that when wealth is included as a continuous variable and endogeneity allowed for, wealth consistently has a clear positive effect.

Most other effects in this model are similar to the previous results, except for some change in province effects, which we suspect may be because these dummies are now capturing the effect of wealth.

3.6.4.3 Model set III: IV Probit Model for Endogenous Wealth Variable (Continuous)

This model is defined as in previous Model sets and the results are given in Table 3.13. Here we allow for a continuous measure of wealth and correct for endogeneity using standard instrumental variables model.

The rates of grade progression estimates in this model are similar to that in the Model set I, except a few differences. Note first though that the continuous household wealth variable shows very similar effect as in the case of the categorical wealth measure used in Model set I. The variable has positive coefficients for all samples, implying for every one unit increment of household wealth a child increases rates of grade progression by 0.063 for entire sample, 0.086 in rural areas, 0.030 in urban areas, 0.055 if a female child and 0.073 if a male child. Note this is after correcting for endogeneity of wealth, confirming our concern about how endogeneity was dealt with in Model set II.

In this model, the dummy variables for the provinces are quite different to those for Model set I, with most provinces having positive and significant coefficients for most samples. In Model set I several provinces did worse than base province (Nairobi). Given the disparity in wealth across provinces, this change is not surprising.

3.6.4.4 Model set IV: Including HIV/AIDS as an Explanatory Variable

The structure of this model set is the same as in previous Model sets except we now include a variable for HIV status of the adults in the household. The results are given in Table 3.16. Sample sizes are much smaller due to the many missing values on this HIV variable.

The variable indicating whether any adult in the household is HIV positive has a significant coefficient of -0.040 for the urban sample with insignificant effects in all other samples. In other words, there appears to be little impact of HIV/AIDS on children's grade progression, except in urban areas.

There are other differences in effects of other variables, but not substantial. Most effects are weaker, reflecting the smaller sample size in this set of estimates.

3.6.4.5 Model set V: Sample Selection Model for Missing Data on HIV Status

The estimates are obtained using IV estimation to deal with endogeneity of wealth, and a sample selection correction is made. Results are in Table 3.20. Results here are very similar to Model set IV. The IMR included to deal with sample selection is not significant, and given this is the only change from Model set IV, it is not surprising that results are very similar.

3.6.4.6 Model set VI: Use of Imputation by Chained Equations for HIV/AIDS Missing Data

This model use IV regression analysis and imputed values for HIV status, and the estimates are given in Table 3.23. This model gives much larger sample size than previous Model sets IV and V. It is this set of results that we present as most reliable, having dealt with wealth endogeneity and minimising loss of observations due to missing values on the HIV variable.

The message from Table 3.23 is quite clear. Males have poorer progression through school grades than females, as do Muslims, compared to other religions. There is some evidence of pre-primary aged children having a negative effect on progression, but there are no other significant sibling effects, either positive or negative. Parental education and wealth have very strong positive effects on progression with wealth having its strongest impact in rural areas. There is some evidence of a negative impact or domestic violence.

Notably, the presence of adults who are HIV positive does not appear to affect progression, nor do other household characteristics such as female circumcision, the practice of polygamy or loss of one or both parents.

3.7 Chapter Discussion and Conclusion

This discussion is based largely on results for Model set VI. This is because the models include all factors including HIV/AIDS and it corrects the possible household wealth endogeneity and sample selection for HIV data using established methods and procedures. However, we will also look into some results from the other Model sets.

The results reveal that the child's age does not play any significant role in the child's school attendance and attainment. The major effect of age is on rates of grade progression. Grade progression is found to decrease up to the age of about 14 years old in the rural areas with a female child doing worse than a male child. It is only after this age that grade progression starts to improve. There are many factors associated to this effect of age in the rural areas. In the rural areas a child is involved in many other activities such as household chores which restrict the child from going to school and being able to concentrate on class work. Also in the rural areas we find less school facilities such as laboratories, libraries and even teachers meaning children are less likely to meet the required standards to go to the next grade.

In the school attendance model, the dummy variable for gender shows a male child much more likely to attend secondary school than a female. This could be a result of parents favouring a male child over the female. This might be the case when there is a restricted household budget where the parents prefer to invest in a male child hoping to get better returns after completion of schooling. A female child is likely to get married and leave the family but a male child will remain in the family and will be the provider and bread winner for the parents when they get old. Despite the efforts parents take to

send their male child to secondary school, rates of grade progression reveal that a male child progresses less compared to a female child during their primary schooling, especially in the rural areas. This indicates that a female might do better than a male if given equal opportunities in their secondary schooling.

Looking into the effect of presence of other siblings we find a mixture of effects of different age groups in the same household. For the primary school level, we find some element of rivalry among primary school aged children in the same household. A female is less likely to attend her primary school if there is at least one primary school aged male in the same household. However, this is not the case if there is a secondary school aged male in the same household where the result show some synergy effect for this male child and a female child. Here a female attends her primary school more if the older male child is in the same household. The presence of a pre-primary school aged male in the same household show a male child is more likely to attend primary school.

In the case of primary school attainment, if in the household there is a female child of the same age, then a male child is more likely to complete his primary school. The results for school attendance suggest that male children are able to achieve a higher grade despite their poorer school attendance. This effect could be because of the parents deciding to educate a male child and not a female child probably due to constrained household budget.

Looking at sibling effects on secondary school attendance, the results show some synergy effect where a female is more likely to attend her secondary school if a primary school aged female is in the same household. This could be as result of sharing household chores where a primary school a child does most of the work, giving the older

female child a chance to attend school. In addition, we find some rivalry effect among the older children where a male child is less likely to attend secondary school if there is at least one secondary school aged female in the same household.

In the case of school attainment, the results show that a female is more likely to complete primary school if there is a pre-primary school aged male in the household. However, a male is found to be more likely to complete primary school if there is a primary school aged female in the same household indicating some sibling synergy effect. This could be due to the female taking on household chores and giving the male time to complete schooling.

In terms of rate of grade progression, we find a female will have lower rates of grade progression if there is a pre-primary school male or female child in the same household.

The findings of "sibling rivalry" are likely a result of households choosing to direct their limited resources to schooling selectively depending on which children are likely to generate better returns to education. Thus, income maximising behaviour on the part of parents, especially in the rural areas plays a big role. With household resource constraints, parents will invest in resources/assets that provide a greater return over their lifetime. This usually takes the form of educating boys (at the cost of educating girls), as often the lifetime return from boys is higher, driven primarily by higher male wage rates. On the other hand, Cornwell *et al.* (2005) explain that there might be positive spillover effects from having children attending school, due to economies of scale in child costs and/or from the development of a "schooling culture" within the household. This spillover essentially results in a form of positive 'externality' in children's educational attainment, which the parents will try to internalize through their schooling decisions on

each subsequent child. There is some evidence of this sibling synergy in the results presented here.

Other evidence supporting the effect of household composition on educational attainment is not straight foreword but remains mixed. For example, the study by Parish and Willis (1993), using data from Taiwan, found that a male with sisters in the household can have much better schooling outcomes. Having older sisters is especially beneficial for younger children as these older sisters are not only more likely to help take care of their younger children, but they are also more likely to take up wage employment (domestic work) that helps pay school fees and allows younger children to postpone entering the labour market early. In the context of Africa, Morduch (2000) finds that in Tanzania moving from a situation where a particular child has all brothers to one where he/she has all sisters, the number of completed years of schooling increases by 0.44 years. He does not however find any inherent gender effects. However, in South Africa, Morduch (2000) finds no statistically significant gender composition effect.

In other Sub-Saharan African countries, during colonial rule and even after independence, girls in Kenya had different educational experiences than boys. During the colonial time, girls in African countries were concerned with preparation of food and other household activities (Staudt, 1985). It was the boys who were singled out for formal education in colonized countries, while the missionaries taught the females domestic skills. This is shown in a study by Duncan (1989) on attitudes and school achievement among students in Botswana, who found strong support among students for the notion that women should be primarily responsible for domestic work. In addition, she established that a gender ideology that defined various subjects as male or female

was a significant factor in determining achievement. One area of crucial concern is women's access to education. Generally, the disproportion between the sexes in education is partly historical (traditions, cultures and customs) and partly economical. Men had a head start in education during the colonial years, and the lead is still continuing. Until recently, rural parents have been unwilling to invest in their daughters' education since they considered such an investment wasteful. In rural families, to send daughters to school would imply a greater cost in terms of foregone contribution to the farm and home. Moreover, given the nature of Kenyan society, once a parent has been paid the bride wealth, he relinquishes all the rights to his daughter and her future earnings will benefit her husband's family, meaning there was no direct economic benefit in educating girls. In our study, the results from attendance do not confirm this argument while results from attainment strongly contradict this argument, suggesting a shift away from these colonial gender attitudes.

In developed countries, some studies have found that children with fewer siblings obtain more schooling than those with more siblings and this negative relationship persists even when family socioeconomic characteristics are controlled for (See for example Goux & Maurin (2005) for France). From developing countries, however, the findings are quite mixed. Evidence from Thailand (Knodel, Havanon, and Sittirai, 1990) and Brazil (Psacharopoulos & Arriagada, 1989) suggest that there is a negative relationship between the number of siblings and educational attainment. In the case of Vietnam (Anh, *et al.,* 1998) the relationship is negative for families with six or more children and the effects are quite small once other family characteristics are controlled for. In their work, Cornwell *et al.* (2005) using South African data, found no gender differences in school attendance and the effect of sibling rivalry is not present in South African

children in school attendance. Evidence from Botswana by Chernichovsky (1985) and from Kenya by Gomes (1984) found that large family size tends to be positively related to educational attainment. Gomes pointed out that parents in Kenya favour their eldest children with educational resources regardless the size of the household. The explanation for this positive relationship typically involves households in Africa drawing on a large kinship network beyond the immediate family, which reduces the costs (financial, emotional and time) associated with additional children. Our results confirm that the presence of other siblings in the same household have mixed effects on schooling. Our results indicate some rivalry and some synergy effects.

In most results we find the death of the child's mother decreases the rate of grade progression, especially for a male child in the rural areas. In some cases, a child having both parents deceased has a large detrimental effect on attendance at primary school. This is expected since the child no longer has anyone to provide educational resources. Children may also be psychologically affected and traumatised, and unable to concentrate in their class work resulting to poor performance and increased grade repetition. In African communities, the father is mostly the sole bread winner and his death affects the entire household income status, and in turn education of the children. The death of the parents means changes in living arrangements, displacements, and lack of availability of resources for schooling, health and food for the children which, as a result it can even contribute to orphaned children becoming homeless. Older children may be forced drop out of school to earn money to provide for their young siblings, or remain home to take care of their young ones or do household chores in place of their deceased parent(s).

Previous literature finds the same effect of orphans on their schooling. Case and Ardington (2006) using longitudinal data from a demographic surveillance area in KwaZulu, South Africa, find maternal orphans being significantly less likely to be enrolled in school and have fewer years of schooling. Barman (2010) in the study in India found Maternal orphans are more disadvantaged than paternal orphans, and double orphan children are really vulnerable in regard to educational attainment.

In this study, the factor of household religion is found to have very strong impact on the child's primary school enrolment. Compared to Catholics, Protestants and Seventh Day Adventists have a much higher primary school enrolment. However, as the child grows up, the choice of denominations does not seem to influence school attendance. Some results suggest primary school completion and grade progression are much lower if a child comes from the Islamic religion. The possible reason could be that Islamic religion is well spread and practiced in the urban areas. In their teaching children go for *"Madrasa"*[36] even during school times which could reduce a child's school attendance and performance in general.

The level of education of the household head is a key determinant of child schooling. We find the child whose parent (household head) has at least primary school level education is more likely to attend his or her primary and secondary school compared to the child whose parent is not educated. The same pattern occurs in grade attainment, where a child is more likely to complete his or her primary school, as well as for grade progression for age. Other literature are in line with our results. For example Barman (2010) suggests that promotion of literacy at the household level may be an important

[36] Madrasa is an Islamic school set up to study pure Islamic religion by the children as they grow.

step for reducing illiteracy. His study finds a positive relationship between parental education, especially mother's education, and educational attainment of children. The results in our study show strongly the positive effect of educated parents on their child's education.

Looking into the household wealth variable we are going to examine the results from Model sets II, III, V and VI where we have taken into account of endogeneity, especially treating wealth as a continuous variable. The results in Model sets III - VI show that as household wealth increases the child is more likely to be attending primary school and improve grade progression. These results show the parent's capacity and willingness to invest in children's education for better a future, or with a view of the child providing for them when they are old and can longer generate their own income. It is worth noting that at the primary school level, wealth does not negatively affect child schooling. This could be because at this stage other factors may play a bigger role. In primary school most children attend public schools which are less expensive and even the poorer households can afford to send their children to schools. At secondary school level, the effect of wealth could be due to the parents' inability to afford fees, resulting to children staying at home until the part or full tuition fees are available. At this time, the child is also more able to engage in child labour so the opportunity cost of schooling is higher. In addition, at this stage we find most of the children are in their teenage years. It is in this stage where they are more likely to engage in other activities and not studies. This could be worsened by the fact a child from a richer family has adequate income to afford drugs, alcohol, etc., which might lead to dropping out from school.

Shavit and Pierce (1991) using data from Israel, find that for the richer Jews, family size has a negative relationship with educational attainment of children, while there is a positive relationship between family size and educational attainment of children in the poor Muslim households or rather poor households. In our study it is shown that wealth is very important for the child's primary school attendance and as increases the child's rates of grade progression. In their study, Loaiza and Lloyd (2008) found that the overwhelming majority of adolescents in Africa are not attending a grade appropriate for their age either because they do not attend school or because they are attending a grade that is behind the grade that is appropriate for their age. With sexual maturation, adolescents face new social challenges in school and are at greater risk than their younger classmates of dropout if they are behind grade for their age. Thus, the educational circumstances of adolescents are in part determined by a critical decision parents make on their children's behalf well before their children become adolescents, and that is the age of first enrolment. These findings are reflected in our study.

Considering the child's residential location, the results from Model set V show that a child from urban areas has lower school attendance than the child from a rural area. One of the possible reasons is that in the cities, education is very expensive as compared to rural areas in terms of cost of living, tuition fee and transport to school. Children from poorer areas, especially from slums, and some middle class households, enrol in day schools which means paying daily transport and lunch at school and this can mean failing to go to school due to a shortage of money. In addition, because of easy access to job opportunities in the urban area compared to rural, this provides an incentive for children to leave school in favour of jobs to raise money for household needs.

Results for Model set V indicate that a child residing in an urban area has better rate of grade progression. This is most likely to do with better quality of school/education provision in the urban areas.

The other factor found to affect child schooling is if any individual in the household is HIV/AIDS positive. In the case of secondary level schooling, as expected, the presence of a HIV positive household member makes a child less likely to attend school. HIV/AIDS also reduces the child school grade progression, particularly in the urban areas. This could be in part a result of the child taking more household responsibilities including taking care of the patient. Considering that the effect is more in the urban areas, a key factor could be the availability of household income. Household income may be diverted towards patient care and treatment in the hospital, at the expense of not paying the school fees for the child. The loss of family resources usually has its biggest impact on the immediate family of the person with HIV/AIDS.

Our study reveals that it is only secondary school aged children whose schooling is affected by HIV/AIDS, and mostly in the urban areas. Mishra *et al*. (2007) show evidence that the child takes most household responsibilities including taking care of the sick person, and diversion of the funds for the treatment of the sick person. Bedi, *et al*. (2004) found that HIV/AIDS did not have any effect on primary school enrolment, further supporting our results. Yamano & Jayne (2005) found a strong correlation between working-age adult mortality and lagged HIV-prevalence rates at nearby sentinel survey sites. They also found school attendance, especially for children in relatively poor households, to have negative correlation with lagged provincial HIV-prevalence rates. Their study found children, especially girls in relatively poor households, are less

likely to be in school directly prior to the death of an adult member than children in unaffected households. By contrast, boys in relatively poor households are less likely to be in school after an adult death. Also, our results shed light on earlier studies by Ainsworth, Beegle and Koda (2002), where based on evidence from six African countries, they reported that countries with higher HIV prevalence appear to have higher enrolment rates, suggesting that differences in educational policies play a greater role than disease incidence in determining outcomes.

Bennell *et al.* (2001), in their study using data from Botswana, found a negative impact of the HIV/AIDS epidemic on education in the country. Hargreaves and Glynn (2002), in their study on educational attainment and HIV-1 infection in developing countries, found that in Africa, higher educational attainment is often associated with a greater risk of HIV infection. Although they indicated that the pattern of new HIV infections might be changing towards a greater burden among less educated groups, overall, the impact will be on the community and in the end the schooling of the children will be badly affected. In addition, a study in rural Zambia indicated about 68% of the orphans of school age were not enrolled in school compared to 48% of non-orphans (UNAIDS, UNICEF, & USAID, 2004). A study of heavily HIV/AIDS affected communities in Zimbabwe, showed that about 48% of primary school aged orphans had dropped out of school, most often at the time of parent's illness or death, and of the children of secondary school age interviewed, there were no orphans who were able to stay in school (UNAIDS, 2000).

The results from Model set V suggest that household violence has a positive effect on the child's primary school attendance and attainment. The reason for this effect is not

clear. Despite this effect, however, the results show a child who experiences household violence will have lower rate of grade of progression during primary schooling. The results also show that a child from a household with violence has low secondary school attendance and grade progression. This could explain why the children are not affected in their primary school attendance and attainment. This could be because the older children remain at home taking care of the young ones, allowing the younger ones to continue in school. In Kenya, mostly household violence include wife beating and when this happens, the wife runs away from the house to their relatives or parents, leaving her family. The children are affected in many ways including psychologically, while the older children especially girls, take over the household activities and household chores.

Several studies have established that young people who flee violent homes are at heightened risk of emotional and behavioural problems (Baker *et al.,* 2002; Edleson, 1999; Miller, 2010) Violence can have a pronounced impact on children's adjustment in school, including their ability to learn. There is a consistent thread running through the research findings of higher levels of aggression, greater likelihood of seeing the intentions of others as hostile, psychosomatic disorders, difficulties with schoolwork, poor academic performance, school phobia and difficulties in concentration and attention (Cummings & Davies, 1994). Our study confirms that household violence affects schooling, in particular for older children in their secondary school enrolment and the rates of grade progression in primary school.

The results from Model set V showed insignificant effect of the variables for households with more than one wife (polygamy) and the female being circumcised. However, the results from the Model sets I – IV and VI showed a polygamous household is likely to

have lower primary school attendance. This is to be expected because many wives means many children from different mothers, mostly living in different houses. This means that the available income or wealth is to be shared among the wives and their children, resulting in insufficient funds to send children to school. Female circumcision is found to reduce primary school attendance. These results are found in the Model set IV. This could be explained by the physiological suffering the female child undergoes before and after circumcision. In many communities circumcision is undertaken at specific times in a year and during this period the children who are circumcised do not go to school until they are healed. In some communities, including Kalenjin and Masai if a girl is circumcised, it means she is engaged to a man chosen by the parents and she ceases to go to school anymore. It is considered to be a mark of beauty and a transition into womanhood and marriage. Alternatively, some girls run away from their homes to institutions (mostly Christian supported) which are against female circumcision, and they stop their schooling for a long period or drop out completely.

To conclude the discussion we need to mention that, apart from the factors included in our models, there are other many potential factors which could be considered if data was available in this kind of study. Such factors may include the distribution of schools and the location of the population. The ratio of teachers to the students is also important. The syllabus and the teaching is supposed to be even in the entire country since students are examined equally at national level, but there are often big variations in quality of education. These and many other factors result in big variations in schooling engagement and outcomes across the country.

After independence, the government had to put effort into providing education to the people. This did not happen well until 2003 when the government implemented a free primary school education policy. Before this time, teachers were responsible for collecting the money from children for their schooling. Many children were forced to drop out of school simply because they could not afford it. Teachers mostly sent children home during the last two weeks of school when the final exams are held, to force parents to pay the fees. Failure to pay fees would mean a child would miss their exams and be forced to repeat the grade.

A study done by Grissom and Shephard, (1989) showed that there is a strong association between progression and high school dropout. Since grade progression is a costly affair, it has been of great interest to understand the policies and practices that might reduce progression (Eide & Showalter, 2001). Cascio (2005) hypothesised that early childhood programs lower the need for progression by promoting school readiness, making progression an outcome of interest in evaluation of public pre-schools. Others have argued that family investments encourage normal school progression, thus linking progression to parental education (Oreopoulos, Page and Stevens, 2003; Page, 2006) and some linking it with number of children in the household (Conley & Glauber, 2005). Resources such as household income and school availability and facilities, including well trained teaching and administrative staff, would play a vital role in child's schooling in general.

Kabubo-Mariara and Mwabu (2007) in their study investigate the determinants of demand for schooling in Kenya. Their results show that child characteristics, parental education and other household characteristics, quality and cost of schooling are

important determinants of demand for education services in Kenya. The results further show that girls would be more affected by policy changes than boys. The findings call for targeting in efforts to boost and sustain demand for schooling in Kenya. The study recommends immediate policy interventions focusing on improving quality of education and poverty alleviation

Increasing educational attainment levels of the current school age population is a particularly important issue. By focussing on the factors that affect the demand for schooling (rather than supply side factors), this chapter identifies several key areas for policy makers to target in their attempt to increase educational attainment and human capital accumulation.

Table 3.2. Descriptive Statistic for Selected Variables

(i). Dependent Variables		Individual sample size	Proportions	Standard Deviation
Primary School Attendance	Dummy = 1 if child currently attends school & aged 6 - 14 omitting above secondary	9515	0.857	0.350
Secondary School Attendance	Dummy= 1 if child currently attends school & aged 15 - 18 omitting above secondary	3340	0.607	0.488
School Grade attainment	Grade attained:			
	0 = No Schooling;	981	0.294	
	1 = Some but not all primary schooling;	1871	0.560	
	2 = Completed Primary School	488	0.146	
Grade progression	grade progression a child has passed to next grade	9281	0.684	0.233

(ii). Individual Characteristics		Independent Variable: Primary School Attendance			Independent Variable: Secondary School Attendance		
		Individual sample size	Proportions	Standard Deviation	Individual sample size	Proportions	Standard Deviation
Male	Dummy = 1 if child is male and aged 0 - 17 years	9515	0.512	0.500	3340	0.515	0.500
Deceased mother	Dummy = 1 if a child in the household's mother is deceased	9515	0.056	0.230	3340	0	0
Deceased father	Dummy = 1 if a child in the household's father is deceased	9515	0.123	0.329	3340	0	0
Both parents deceased	Dummy = 1 if both parents of a child in the household are deceased	9515	0.029	0.168	3340	0	0
Female circumcision	Dummy = 1 if the female child is circumcised	8291	0.405	0.478	2897	0.331	0.471
(iii). Household Characteristics							
	Religion						
Catholic	Dummy = 1 if the household belong to Catholic religion**(BASE)**	8303	0.225	0.417	2904	0.242	0.428
Protestant & other Christianity	Dummy = 1 if the household belong to Protestant/Other Christian religion	8303	0.573	0.500	2904	0.597	0.490
Muslim	Dummy = 1 if the household belong to Muslim religion	8303	0.176	0.360	2904	0.138	0.345

Table 3.2: ***Continued***

		Independent Variable: Primary School Attendance			Independent Variable: Secondary School Attendance		
		Individual sample size	**Proportions**	**Standard Deviation**	**Individual sample size**	**Proportions**	**Standard Deviation**
	Wealth Index						
Poorest household	Dummy = 1 if the household belong is classified as poorest (in terms of wealth) **(BASE)**	9515	0.251	0.434	3340	0.193	0.395
Poorer household	Dummy = 1 if the household belong is classified as poorer (in terms of wealth)	9515	0.194	0.396	3340	0.190	0.392
Middle household	Dummy = 1 if the household belong is classified as middle (in terms of wealth)	9515	0.199	0.399	3340	0.204	0.403
Richer household	Dummy = 1 if the household belong is classified as richer (in terms of wealth)	9515	0.187	0.390	3340	0.201	0.402
Richest household	Dummy = 1 if the household belong is classified as richest (in terms of wealth)	9515	0.168	0.374	3340	0.211	0.408
	Residence/Region						
Urban	Dummy = 1 if the household resides in urban area	9515	0.217	0.412	3340	0.256	0.436
Nairobi	Dummy = 1 if the household resides in Nairobi Province **(BASE)**	9515	0.063	0.243	3340	0.086	0.281
Central	Dummy = 1 if the household resides in Central Province	9515	0.145	0.352	3340	0.156	0.363
Coast	Dummy = 1 if the household resides in Coast Province	9515	0.113	0.317	3340	0.112	0.316
Eastern	Dummy = 1 if the household resides in Eastern Province	9515	0.130	0.336	3340	0.126	0.332
Nyanza	Dummy = 1 if the household resides in Nyanza Province	9515	0.129	0.335	3340	0.148	0.355
Rift Valley	Dummy = 1 if the household resides in Rift Valley Province	9515	0.193	0.395	3340	0.164	0.371
Western	Dummy = 1 if the household resides in Western Province	9515	0.125	0.331	3340	0.139	0.346
North Eastern	Dummy = 1 if the household resides in North Eastern Province	9515	0.102	0.302	3340	0.069	0.253
	Others Variables						
Head education	Dummy = 1 if the household head has some primary school education	9515	0.721	0.448	3340	0.895	0.307
Violence	Dummy = 1 if the wife of the household experienced any form of violence	8300	0.250	0.413	2905	0.142	0.349
Any HIV patient	Dummy = 1 if there is any HIV positive member in the household	4003	0.102	0.303	1402	0.121	0.326
Any agree HIV test	Dummy = 1 if any household member agree to be tested for HIV/AIDS, given was selected	4525	0.885	0.319	1588	0.883	0.322

Table 3.2: ***Continued***

		Independent Variable: Primary School Attendance			Independent Variable: Secondary School Attendance		
		Individual sample size	**Proportions**	**Standard Deviation**	**Individual sample size**	**Proportions**	**Standard Deviation**
Imputed HIV patient	Dummy = 1 if there is any HIV positive member in the household after data imputation	8636	0.107	0.310	3060	0.112	0.316
Polygamy	Dummy = 1 if the household is polygamous	8299	0.127	0.333	2905	0.073	0.260

Model Set I: All independent variables (including wealth) are assumed to be exogenous.

No HIV/AIDS variable included.

Table 3.3. Probit Estimates for School Enrolment using Age-Qualification

3.3a) Dependent Variable: *Primary School Attendance*			**3.3b) Dependent Variable: *Secondary School Attendance***	
	Number of observations = 8,282		Number of observation = 2,895	
	Wald χ^2 = 1,153.39		Wald χ^2 = 604.22	
	Prob > χ^2 = 0.000		Prob > χ^2 = 0.000	
	Degrees of freedom = 38		Degrees of freedom = 35	
	Coefficient	**Marginal effect**	**Coefficient**	**Marginal effect**
Age	0.103	0.007	0.643	0.248
	(0.091)	(0.006)	(0.927)	(0.358)
Age squared	-0.016***	-0.001***	-0.032	-0.012
	(0.005)	(0.000)	(0.028)	(0.011)
Male	0.153	0.010	0.968***	0.357***
	(0.135)	(0.009)	(0.190)	(0.064)
Male pre-primary aged siblings	-0.012	-0.001	-0.200***	-0.077***
	(0.045)	(0.003)	(0.060)	(0.023)
Male x Male pre-primary aged siblings	-0.076	-0.005	0.196**	0.076**
	(0.059)	(0.004)	(0.084)	(0.032)
Male primary school aged siblings	-0.064*	-0.004*	0.162***	0.063***
	(0.035)	(0.002)	(0.045)	(0.017)
Male x Male primary school aged siblings	0.058	0.004	-0.142**	-0.055**
	(0.052)	(0.003)	(0.061)	(0.024)
Male secondary school aged siblings	0.135**	0.009**	0.113	0.043
	(0.062)	(0.004)	(0.076)	(0.029)
Male x Male secondary school aged siblings	-0.045	-0.003	-0.149	-0.058
	(0.089)	(0.006)	(0.119)	(0.046)
Female pre-primary aged siblings	0.008	0.001	-0.153**	-0.059**
	(0.048)	(0.003)	(0.060)	(0.023)
Male x Female pre-primary aged siblings	0.012	0.001	0.257***	0.099***
	(0.066)	(0.004)	(0.091)	(0.035)
Female primary school aged siblings	-0.036	-0.002	0.159***	0.062***
	(0.037)	(0.002)	(0.051)	(0.020)
Male x Female primary school aged siblings	0.008	0.001	-0.039	-0.015
	(0.050)	(0.003)	(0.070)	(0.027)
Female secondary school aged siblings	0.095	0.006	0.281***	0.108***
	(0.089)	(0.006)	(0.088)	(0.034)
Male x Female secondary school aged siblings	-0.028	-0.002	-0.527***	-0.203***
	(0.107)	(0.007)	(0.122)	(0.047)
Deceased mother	0.010	0.001		
	(0.161)	(0.010)		
Deceased father	0.059	0.004		
	(0.108)	(0.007)		
Both parents deceased	-0.678***	-0.081***		
	(0.243)	(0.045)		
Protestant & other Christianity	0.387***	0.027***	-0.015	-0.006
	(0.071)	(0.006)	(0.067)	(0.026)
Muslim	-0.033	-0.002	-0.132	-0.052
	(0.124)	(0.009)	(0.163)	(0.064)

Table 3.3: ***Continued***

	3.3a) Dependent Variable: ***Primary School Attendance***		**3.3b) Dependent Variable:** ***Secondary School Attendance***	
	Coefficient	**Marginal effect**	**Coefficient**	**Marginal effect**
Poorer household	0.232***	0.013***	-0.045	-0.017
	(0.078)	(0.004)	(0.102)	(0.040)
Middle household	0.415***	0.022***	-0.079	-0.031
	(0.088)	(0.004)	(0.103)	(0.040)
Richer household	0.419***	0.022***	-0.162	-0.063
	(0.108)	(0.005)	(0.104)	(0.041)
Richest household	0.526***	0.026***	-0.356**	-0.140**
	(0.160)	(0.006)	(0.141)	(0.056)
Head education	2.451***	0.443***	2.903***	0.702***
	(0.102)	(0.020)	(0.234)	(0.013)
Urban	0.014	0.001	-0.207*	-0.081*
	(0.108)	(0.007)	(0.107)	(0.042)
Central	0.384**	0.020**	-0.056	-0.022
	(0.209)	(0.008)	(0.143)	(0.056)
Coast	0.314**	0.017**	-0.121	-0.047
	(0.188)	(0.008)	(0.145)	(0.057)
Eastern	0.504**	0.024**	0.052	0.020
	(0.208)	(0.007)	(0.149)	(0.057)
Nyanza	0.802***	0.032***	0.204	0.077
	(0.202)	(0.005)	(0.143)	(0.053)
Rift Valley	0.182	0.011	0.033	0.013
	(0.189)	(0.010)	(0.145)	(0.055)
Western	0.581***	0.026***	0.356***	0.131***
	(0.203)	(0.006)	(0.147)	(0.051)
North Eastern	-0.383*	-0.034	0.213	0.080
	(0.207)	(0.024)	(0.242)	(0.088)
Violence	-0.024	-0.002	-0.163*	-0.064*
	(0.070)	(0.005)	(0.093)	(0.037)
Polygamy	-0.291***	-0.024***	-0.071	-0.028
	(0.078)	(0.008)	(0.121)	(0.047)
Female circumcision	-0.131*	-0.009*	-0.066	-0.026
	(0.073)	(0.005)	(0.075)	(0.029)
constant	-0.053		-4.606	
	(0.460)		(7.656)	

*significant at 10%; **significant at 5%; ***significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)
Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 3.4. Ordered Probit Estimates for School Attainment using Age-Qualification

Number of obs = 2,895	Prob > χ^2 = 0.0000	**Marginal effects after Ordered Probit**		
Wald χ^2 = 332.52	Degrees of freedom = 35	**No schooling**	**Some Primary**	**Complete Primary**
Dependent Variable: *School Grade Attainment*	**Coefficient**	**Marginal effect**	**Marginal Effect**	**Marginal Effect**
Age	-0.831	0.298	-0.163	-0.135
	(0.729)	(0.261)	(0.143)	(0.119)
Age squared	0.023	-0.008	0.005	0.004
	(0.022)	(0.008)	(0.004)	(0.004)
Male	-0.469***	0.168***	-0.093***	-0.075***
	(0.135)	(0.048)	(0.028)	(0.022)
Male pre-primary aged siblings	0.071	-0.025	0.014	0.012
	(0.054)	(0.019)	(0.011)	(0.009)
Male x Male pre-primary aged siblings	-0.032	0.011	-0.006	-0.005
	(0.067)	0.024)	(0.013)	(0.011)
Male primary school aged siblings	-0.074**	0.027**	-0.014**	-0.012**
	(0.032)	(0.012)	(0.006)	(0.005)
Male x Male primary school aged siblings	0.115***	-0.041***	0.022***	0.019***
	(0.043)	(0.015)	(0.008)	(0.007)
Male secondary school aged siblings	-0.066	0.024	-0.013	-0.011
	(0.057)	(0.020)	(0.011)	(0.009)
Male x Male secondary school aged siblings	-0.038	0.014	-0.007	-0.006
	(0.080)	(0.029)	(0.016)	(0.013)
Female pre-primary aged siblings	0.086*	-0.031*	0.017*	0.014*
	(0.044)	(0.016)	(0.009)	(0.007)
Male x Female pre-primary aged siblings	-0.075	0.027	-0.015	-0.012
	(0.061)	(0.022)	(0.012)	(0.010)
Female primary school aged siblings	-0.056	0.020	-0.011	-0.009
	(0.037)	(0.013)	(0.007)	(0.006)
Male x Female primary school aged siblings	0.074	-0.027	0.015	0.012
	(0.049)	(0.018)	(0.010)	(0.008)
Female secondary school aged siblings	-0.187***	0.067***	-0.037***	-0.030***
	(0.060)	(0.021)	(0.012)	(0.010)
Male x Female secondary school aged siblings	0.309***	-0.111***	0.061***	0.050***
	(0.089)	(0.032)	(0.018)	(0.014)
Protestant & other Christianity	-0.027	0.010	-0.005	-0.004
	(0.052)	(0.019)	(0.010)	(0.009)
Muslim	0.014	-0.005	0.003	0.002
	(0.125)	(0.045)	(0.024)	(0.021)
Poorer household	0.062	-0.022	0.012	0.010
	(0.057)	(0.020)	(0.010)	(0.010)
Middle household	0.040	-0.014	0.008	0.007
	(0.063)	(0.022)	(0.012)	(0.011)
Richer household	-0.091	0.033	-0.019	-0.014
	(0.068)	(0.025)	(0.015)	(0.010)
Richest Household	-0.153	0.056	-0.033	-0.023
	(0.114)	(0.043)	(0.027)	(0.016)
Head Education	3.378***	-0.786***	0.626***	0.161***
	(0.283)	(0.009)	(0.011)	(0.008)

Table 3.4: ***Continued***

		Marginal effects after Ordered Probit		
		No schooling	**Some Primary**	**Complete Primary**
Dependent Variable: ***School Grade Attainment***	**Coefficient**	**Marginal effect**	**Marginal Effect**	**Marginal Effect**
Urban	0.077	-0.027	0.014	0.013
	(0.089)	(0.031)	(0.016)	(0.015)
Central	-0.176	0.065	-0.038	-0.026
	(0.150)	(0.056)	(0.036)	(0.021)
Coast	0.065	-0.023	0.012	0.011
	(0.151)	(0.053)	(0.026)	(0.026)
Eastern	0.064	-0.023	0.012	0.011
	(0.145)	(0.051)	(0.026)	(0.025)
Nyanza	-0.342**	0.128**	-0.081**	-0.047***
	(0.141)	(0.055)	(0.039)	(0.016)
Rift Valley	0.044	-0.016	0.008	0.007
	(0.143)	(0.050)	(0.026)	(0.024)
Western	-0.144	0.053	-0.031	-0.022
	(0.144)	(0.054)	(0.034)	(0.020)
North Eastern	-0.317*	0.119	-0.077	-0.043**
	(0.189)	(0.074)	(0.053)	(0.021)
Violence	0.121*	-0.043*	0.022*	0.021 *
	(0.068)	(0.023)	(0.011)	(0.012)
Polygamy	0.006	-0.002	0.001	0.001
	(0.076)	(0.027)	(0.015)	(0.012)
Female circumcision	0.064	-0.023	0.012	0.010
	(0.056)	(0.020)	(0.011)	(0.009)
τ_{1E}	-5.078			
	(5.972)			
τ_{2E}	-3.275			
	(5.973)			

*significant at 10%; **significant at 5%; ***significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)
Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 3.5. Regression Estimates for Rate of Progression during Primary and Secondary School using Age-Qualification

Dependent Variable: ***Rate of Progression***	**A: Overall** Number of observations = 8,109 F(38, 3,733) = 69.91	**B: Rural** Number of observations = 6,288 F(36, 2,794) = 53.78	**C: Urban** Number of observations = 1,821 F(37, 938) = 14.53	**D: Female** Number of observations = 3,945 F(31, 2,579) = 41.32	**E: Male** Number of observations = 4,164 F(31, 2,675) = 52.51
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
Age	-0.124***	-0.128***	-0.112***	-0.115***	-0.133***
	(0.005)	(0.006)	(0.011)	(0.007)	(0.008)
Age squared	0.005***	0.005***	0.004***	0.004***	0.005***
	(0.000)	(0.000)	(I0.000)	(0.000)	(0.000)
Male	-0.032***	-0.048***	-0.019		
	(0.012)	(0.014)	(0.022)		
Male pre-primary aged siblings	-0.027***	-0.028***	-0.039***	-0.030***	-0.031***
	(0.005)	(0.006)	(0.012)	(0.005)	(0.005)
Male x Male pre-primary aged siblings	-0.005	-0.002	0.002		
	(0.006)	(0.007)	(0.014)		
Male primary school aged siblings	-0.007*	-0.004	-0.024**	-0.008*	0.000
	(0.004)	(0.005)	(0.010)	(0.004)	(0.005)
Male x Male primary school aged siblings	0.007	0.003	0.032***		
	(0.005)	(0.006)	(0.011)		
Male secondary school aged siblings	-0.006	-0.001	-0.033**	-0.007	-0.009
	(0.006)	(0.007)	(0.014)	(0.006)	(0.007)
Male x Male secondary school aged siblings	-0.003	-0.007	0.034**		
	(0.008)	(0.008)	(0.017)		
Female pre-primary aged siblings	-0.012**	-0.018***	0.006	-0.014***	-0.005
	(0.005)	(0.006)	(0.011)	(0.005)	(0.006)
Male x Female pre-primary aged siblings	0.006	0.012*	-0.010		
	(0.006)	(0.007)	(0.013)		
Female primary school aged siblings	0.004	0.003	0.006	0.003	0.001
	(0.004)	(0.004)	(0.009)	(0.004)	(0.004)
Male x Female primary school aged siblings	-0.004	-0.000	-0.010		
	(0.005)	(0.005)	(0.011)		
Female secondary school aged siblings	-0.005	-0.012	0.006	-0.003	-0.004
	(0.006)	(0.008)	(0.008)	(0.006)	(0.009)
Male x Female secondary school aged siblings	-0.000	0.008	-0.016		
	(0.008)	(0.009)	(0.015)		

Table 3.5: ***Continued***

Dependent Variable: *Rate of Progression*	A: Overall	B: Rural	C: Urban	D: Female	E: Male
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
Deceased mother	-0.016	-0.008	-0.046	-0.004	-0.028
	(0.020)	(0.025)	(0.034)	(0.029)	(0.029)
Deceased father	-0.023**	-0.039***	0.026	-0.025	-0.019
	(0.012)	(0.013)	(0.023)	(0.016)	(0.015)
Both parents deceased	0.015	0.020	-0.003	-0.000	0.029
	(0.029)	(0.035)	(0.051)	(0.040)	(0.041)
Protestant & other Christianity	0.007	0.009	0.003	0.012	0.002
	(0.007)	(0.007)	(0.013)	(0.008)	(0.009)
Muslim	-0.039***	-0.002	-0.098***	-0.035*	-0.044**
	(0.015)	(0.019)	(0.023)	(0.020)	(0.019)
Poorer household	0.047***	0.047***	0.027	0.044***	0.049***
	(0.010)	(0.010)	(0.052)	(0.013)	(0.012)
Middle household	0.061***	0.059***	0.044	0.063***	0.060***
	(0.010)	(0.010)	(0.046)	(0.013)	(0.013)
Richer household	0.126***	0.131***	0.071*	0.122***	0.130***
	(0.010)	(0.010)	(0.041)	(0.013)	(0.012)
Richest household	0.179***	0.171***	0.134***	0.172***	0.187***
	(0.013)	(0.016)	(0.041)	(0.017)	(0.017)
Head education	0.191***	0.182***	0.217***	0.174***	0.203***
	(0.013)	(0.014)	(0.026)	(0.019)	(0.016)
Urban	0.022**			0.012	0.032**
	(0.011)			(0.014)	(0.013)
Central	0.045***	0.132***	0.001	0.058***	0.030*
	(0.014)	(0.017)	(0.024)	(0.018)	(0.018)
Coast	-0.054***		-0.011	-0.030	-0.077***
	(0.015)		(0.018)	(0.020)	(0.019)
Eastern	-0.029**	0.050***	0.020	-0.014	-0.047***
	(0.014)	(0.017)	(0.028)	(0.019)	(0.018)
Nyanza	-0.013	0.074***	-0.026	-0.004	-0.025
	(0.014)	(0.018)	(0.020)	(0.018)	(0.018)
Rift Valley	-0.024*	0.062***	-0.049**	-0.025	-0.025
	(0.014)	(0.017)	(0.023)	(0.018)	(0.018)
Western	-0.038***	0.046***	-0.040*	-0.020	-0.058***
	(0.014)	(0.017)	(0.023)	(0.019)	(0.018)
North Eastern	-0.006	0.049*	0.004	-0.032	0.002
	(0.023)	(0.025)	(0.037)	(0.035)	(0.028)
Violence	-0.026***	-0.024***	-0.038**	-0.037***	-0.015

Table 3.5: ***Continued***

Dependent Variable: ***Rate of Progression***	**A: Overall**	**B: Rural**	**C: Urban**	**D: Female**	**E: Male**
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
	(0.008)	(0.009)	(0.018)	(0.011)	(0.010)
Polygamy	-0.015	-0.013	-0.023	-0.011	-0.019
	(0.011)	(0.012)	(0.025)	(0.015)	(0.013)
Female circumcision	-0.007	-0.008	-0.007	0.001	-0.014
	(0.007)	(0.008)	(0.016)	(0.009)	(0.009)
Constant	1.241***	1.194***	1.208***	1.200***	1.254***
	(0.034)	(0.040)	(0.070)	(0.045)	(0.047)

*significant at 10%; **significant at 5%; ***significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)

Model Set II: A case of Endogenous Wealth Variable.[37] No HIV/AIDS variable included.

Table 3.6. **Bivariate Ordered Probit Estimates for School Enrolment using Age-Qualification**

	3.6a) Dependent Variable: *Primary School Attendance*		**3.6b) Dependent Variable: *Secondary School Attendance***	
	Number of observations	= 8,270	Number of observation	= 2,894
	Wald χ^2	= 1,185.94	Wald χ^2	= 694.29
	Prob > χ^2	= 0.000	Prob > χ^2	= 0.000
	Degrees of freedom	= 38	Degrees of freedom	= 35
	Coefficient	**Marginal Effect**	**Coefficient**	**Marginal effect**
Age	0.138	0.009	0.802	0.311
	(0.093)	(0.006)	(0.914)	(0.354)
Age squared	-0.017***	-0.001***	-0.036	-0.014
	(0.005)	(0.000)	(0.028)	(0.011)
Male	0.150	0.010	0.937***	0.348***
	(0.134)	(0.009)	(0.184)	(0.063)
Male pre-primary aged siblings	0.007	0.000	-0.207***	-0.08***
	(0.045)	(0.003)	(0.057)	(0.022)
Male x Male pre-primary aged siblings	-0.080	-0.005	0.172**	0.067**
	(0.058)	(0.004)	(0.082)	(0.032)
Male primary school aged siblings	-0.058*	-0.004*	0.142***	0.055***
	(0.035)	(0.002)	(0.045)	(0.017)
Male x Male primary school aged siblings	0.059	0.004	-0.137**	-0.053**
	(0.052)	(0.004)	(0.059)	(0.023)
Male secondary school aged siblings	0.142**	0.010**	0.105	0.041
	(0.062)	(0.004)	(0.075)	(0.029)
Male x Male secondary school aged siblings	-0.042	-0.003	-0.137	-0.053
	(0.089)	(0.006)	(0.114)	(0.044)
Female pre-primary aged siblings	0.023	0.002	-0.174***	-0.067***
	(0.050)	(0.003)	(0.058)	(0.023)
Male x Female pre-primary aged siblings	0.009	0.001	0.248***	0.096***
	(0.066)	(0.004)	(0.088)	(0.034)
Female primary school aged siblings	-0.032	-0.002	0.143***	0.055***
	(0.037)	(0.003)	(0.049)	(0.019)
Male x Female primary school aged siblings	0.008	0.001	-0.042	-0.016
	(0.050)	(0.003)	(0.067)	(0.026)
Female secondary school aged siblings	0.096	0.007	0.275***	0.106***
	(0.089)	(0.006)	(0.087)	(0.033)
Male x Female secondary school aged siblings	-0.023	-0.002	-0.506***	-0.196***
	(0.107)	(0.007)	(0.119)	(0.046)
Deceased mother	-0.006	-0.000		
	(0.159)	(0.011)		
Deceased father	0.065	0.004		
	(0.107)	(0.007)		
Both parents deceased	-0.675***	-0.082***		
	(0.241)	(0.045)		

[37] Wealth is an endogenous ordinal variable with five (5) indexes. We to use bivariate ordered probit model (New Stata Command). We have given variable estimates for the second wealth model. For the case of child's grade progression, we have used the predicted probabilities of wealth variable in the main model.

Table 3.6: ***Continued***

	3.6a) Dependent Variable: *Primary School Attendance*		3.6b) Dependent Variable: *Secondary School Attendance*	
	Coefficient	**Marginal Effect**	**Coefficient**	**Marginal effect**
Protestant & other Christianity	0.369***	0.027***	0.015	0.006
	(0.072)	(0.006)	(0.067)	(0.026)
Muslim	-0.047	-0.003	-0.122	-0.048
	(0.125)	(0.009)	(0.153)	(0.060)
Poorer household	0.394***	0.022***	-0.329**	-0.129**
	(0.121)	(0.006)	(0.144)	(0.057)
Middle household	0.688***	0.033***	-0.565***	-0.221***
	(0.182)	(0.008)	(0.198)	(0.077)
Richer household	0.783***	0.035***	-0.860***	-0.332***
	(0.241)	(0.009)	(0.270)	(0.098)
Richest household	1.067***	0.042***	-1.376***	-0.503***
	(0.353)	(0.011)	(0.390)	(0.114)
Head education	2.364***	0.424***	2.923***	0.699***
	(0.114)	(0.022)	(0.228)	(0.013)
Urban	-0.215	-0.016	0.230	0.088
	(0.167)	(0.015)	(0.195)	(0.073)
Central	0.412**	0.022**	-0.084	-0.033
	(0.208)	(0.008)	(0.141)	(0.055)
Coast	0.452**	0.022**	-0.304*	-0.120*
	(0.199)	(0.008)	(0.156)	(0.062)
Eastern	0.589***	0.027***	-0.078	-0.030
	(0.212)	(0.007)	(0.155)	(0.061)
Nyanza	0.909***	0.035***	-0.011	-0.004
	(0.208)	(0.006)	(0.161)	(0.062)
Rift Valley	0.297	0.017	-0.133	-0.052
	(0.200)	(0.010)	(0.153)	(0.060)
Western	0.702***	0.030***	0.134	0.051
	(0.211)	(0.006)	(0.170)	(0.064)
North Eastern	-0.144	-0.011	-0.209	-0.082
	(0.249)	(0.021)	(0.291)	(0.116)
Violence	-0.006	-0.000	-0.179*	-0.070*
	(0.069)	(0.005)	(0.092)	(0.036)
Polygamy	-0.276***	-0.023***	-0.112	-0.044
	(0.078)	(0.008)	(0.122)	(0.048)
Female circumcision	-0.119	-0.008	-0.091	-0.035
	(0.073)	(0.005)	(0.073)	(0.028)
τ_{1W}	-3.890		2.776	
	(0.398)		(5.601)	
τ_{2W}	-3.131		3.575	
	(0.397)		(5.602)	
τ_{3W}	-2.377		4.327	
	(0.397)		(5.602)	
τ_{4W}	-1.296		5.441	
	(0.396)		(5.603)	
ρ	-0.163		0.311	
	(0.092)		(0.109)	
	Wald test of independent equations **(ρ = 0)**:		Wald test of independent equations **(ρ = 0)**:	
	$\chi_1^2 = 3.01$	$\Pr > \chi^2 = 0.083$	$\chi_1^2 = 7.18$	$\Pr > \chi^2 = 0.007$

Table 3.7. **Bivariate Ordered Probit for School Enrolment:** ***Equation for Household Wealth***

Dependent Variable: ***Household Wealth(Ordinal)***		
	Dependent Variable: ***Primary School Attendance***	**Dependent Variable:** ***Secondary School Attendance***
	Coefficient	**Coefficient**
Excluded Instruments		
Professional, technical or Managerial	0.758***	0.824***
	(0.082)	(0.113)
Clerical	0.712***	0.587***
	(0.168)	(0.169)
Sales	0.207***	0.177**
	(0.060)	(0.081)
Agricultural - self employed	-0.226***	-0.123
	(0.059)	(0.080)
Household & Domestic	0.498***	0.672***
	(0.126)	(0.144)
Services	0.348***	0.232**
	(0.091)	(0.116)
Skilled Manual	0.169	0.444***
	(0.115)	(0.159)
Unskilled Manual	0.164***	0.134
	(0.061)	(0.085)
Embu	0.842***	0.838**
	(0.288)	(0.375)
Kalenjin	-0.548***	-0.375
	(0.206)	(0.301)
Kamba	0.198	-0.006
	(0.245)	(0.335)
Kikuyu	0.356*	0.405
	(0.207)	(0.292)
Kisii	0.300	0.795**
	(0.236)	(0.334)
Luhya	0.133	0.506*
	(0.193)	(0.280)
Luo	0.526**	1.015***
	(0.226)	(0.318)
Masai	-0.895***	-0.476
	(0.228)	(0.361)
Meru	0.710***	0.637*
	(0.252)	(0.343)
Mijikenda/Swahili	-0.750***	-0.597*
	(0.244)	(0.317)
Somali	-0.262	-0.071
	(0.280)	(0.372)
Taita	-0.426	-0.197
	(0.264)	(0.334)
Turkana	-0.713***	-0.709*
	(0.266)	(0.382)
Kuria	0.165	0.341
	(0.317)	(0.434)

Table 3.7: ***Continued***

Dependent Variable: ***Household Wealth(Ordinal)***

	Primary School Attendance	***Secondary School Attendance***
	Coefficient	**Coefficient**
Others		
Age	-0.167***	0.607
	(0.040)	(0.679)
Age squared	0.007***	-0.017
	(0.002)	(0.021)
Household head age	-0.005	-0.006
	(0.010)	(0.011)
Household head age squared	-0.000	-0.000
	(0.000)	(0.000)
Number of children aged 0 and 5 years	-0.108***	-0.106***
	(0.023)	(0.029)
Number of children aged 6 and 14 years	-0.013	-0.047**
	(0.021)	(0.022)
Number of children aged 15 and 18 years	-0.013	0.038
	(0.032)	(0.052)
Protestant & other Christianity	0.033	0.106*
	(0.047)	(0.060)
Muslim	0.180	0.258
	(0.134)	(0.161)
Head education	0.361***	0.371***
	(0.045)	(0.105)
Urban	1.508***	1.521***
	(0.079)	(0.100)
Central	-1.705***	-1.470***
	(0.249)	(0.266)
Coast	-1.645***	-1.369***
	(0.251)	(0.291)
Eastern	-2.170***	-1.708***
	(0.263)	(0.328)
Nyanza	-2.437***	-2.606***
	(0.269)	(0.306)
Rift Valley	-1.550***	-1.504***
	(0.237)	(0.258)
Western	-2.207***	-2.225***
	(0.241)	(0.271)
North Eastern	-3.081***	-2.970***
	(0.263)	(0.336)
Violence	-0.057	-0.152**
	(0.052)	(0.077)
Polygamy	-0.020	-0.075
	(0.071)	(0.102)
Constant	-0.165*	0.322***
	(0.095)	(0.120)

*significant at 10%; **significant at 5%; ***significant at 1%

Robust standard errors are in parentheses (adjusted for clustering on household level)

Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 3.8. **Bivariate Ordered Probit Estimates for School Attainment using Age-Qualification**

Number of obs = 2,894	$prob > \chi^2$ = 0.0000	**Marginal effects after Ordered Probit**		
Wald χ^2 = 370.40	Degrees of freedom = 35	**No schooling**	**Some Primary**	**Complete Primary**
Dependent Variable: *School Grade attainment*	**Coefficient**	**Marginal effect**	**Marginal Effect**	**Marginal Effect**
Age	-0.599	0.217	-0.114	-0.102
	(0.753)	(0.271)	(0.146)	(0.126)
Age squared	0.017	-0.006	0.003	0.003
	(0.023)	(0.008)	(0.004)	(0.004)
Male	-0.431***	0.156***	-0.084***	-0.072***
	(0.144)	(0.051)	(0.031)	(0.022)
Male pre-primary aged siblings	0.049	-0.018	0.009	0.008
	(0.059)	(0.021)	(0.012)	(0.010)
Male x Male pre-primary aged siblings	-0.040	0.014	-0.008	-0.007
	(0.064)	(0.023)	(0.012)	(0.011)
Male primary school aged siblings	-0.084**	0.030**	-0.016***	-0.014**
	(0.033)	(0.012)	(0.006)	(0.006)
Male x Male primary school aged siblings	0.107**	-0.039**	0.020**	0.018**
	(0.043)	(0.015)	(0.009)	(0.007)
Male secondary school aged siblings	-0.063	0.023	-0.012	-0.011
	(0.056)	(0.020)	(0.011)	(0.010)
Male x Male secondary school aged siblings	-0.039	0.014	-0.007	-0.007
	(0.078)	(0.028)	(0.015)	(0.013)
Female pre-primary aged siblings	0.057*	-0.021	0.011	0.010
	(0.051)	(0.018)	(0.010)	(0.008)
Male x Female pre-primary aged siblings	-0.071	0.026	-0.013	-0.012
	(0.060)	(0.022)	(0.012)	(0.010)
Female primary school aged siblings	-0.062***	0.022*	-0.019	-0.011
	(0.037)	(0.013)	(0.007)	(0.007)
Male x Female primary school aged siblings	0.067***	-0.024	0.013*	0.011
	(0.049)	(0.018)	(0.010)	(0.008)
Female secondary school aged siblings	-0.175	0.063***	-0.033***	-0.030***
	(0.059)	(0.021)	(0.012)	(0.010)
Male x Female secondary school aged siblings	0.294	-0.106***	0.056***	0.050***
	(0.090)	(0.032)	(0.019)	(0.015)
Protestant & other Christianity	-0.002	0.001	-0.000	-0.000
	(0.057)	(0.021)	(0.011)	(0.010)
Muslim	0.020	-0.007	0.004	0.004
	(0.130)	(0.047)	(0.024)	(0.023)
Poorer household	-0.202	0.075	-0.043	-0.032
	(0.261)	(0.099)	(0.058)	(0.041)
Middle household	-0.421	0.158	-0.098	-0.061
	(0.448)	(0.175)	(0.114)	(0.062)
Richer household	-0.760	0.289	-0.192	-0.097
	(0.646)	(0.248)	(0.176)	(0.072)
Richest household	-1.143	0.430	-0.301	-0.129
	(0.982)	(0.344)	(0.257)	(0.088)
Head education	3.410***	-0.782***	0.611***	0.171***
	(0.300)	(0.016)	(0.033)	(0.019)
Urban	0.491	-0.166)	0.068**	0.098
	(0.422)	(0.133)	(0.028)	(0.106)
Central	-0.201	0.075	-0.043	-0.031
	(0.151)	(0.058)	(0.036)	(0.022)
Coast	-0.117	0.043	-0.024	-0.019
	(0.240)	(0.090)	(0.052)	(0.038)

Table 3.8: ***Continued***

		Marginal effects after Ordered Probit		
		No schooling	**Some Primary**	**Complete Primary**
Dependent Variable: ***School Grade attainment***	**Coefficient**	**Marginal effect**	**Marginal Effect**	**Marginal Effect**
Eastern	-0.069	0.025	-0.014	-0.011
	(0.204)	(0.075)	(0.042)	(0.033)
Nyanza	-0.530**	0.202**	-0.132**	-0.071**
	(0.231)	(0.092)	(0.062)	(0.031)
Rift Valley	-0.119	0.044	-0.024	-0.019
	(0.228)	(0.085)	(0.049)	(0.037)
Western	-0.345	0.130	-0.080	-0.050
	(0.246)	(0.097)	(0.062)	(0.035)
North Eastern	-0.738	0.286	-0.204	-0.082**
	(0.455)	(0.176)	(0.138)	(0.039)
Violence	0.092	-0.0329	0.016	0.016
	(0.077)	(0.027)	(0.013)	(0.014)
Polygamy	-0.030	0.011	-0.006	-0.005
	(0.085)	(0.031)	(0.017)	(0.014)
Female circumcision	0.043	-0.016	0.008	0.007
	(0.058)	(0.021)	(0.011)	(0.010)
τ_{1E}	-3.580			
	(6.042)			
τ_{2E}	-1.834			
	(6.021)			
ρ	0.299			
	(0.285)			
Wald test of independent equations: ($\rho = 0$): $\chi_1^2 = 0.97$			$\Pr > \chi^2 = 0.324$	

Table 3.9. **Bivariate Ordered Probit for School Attainment:** ***Equation for Household Wealth***

Dependent Variable: ***Household Wealth (Ordinal)***	
	School Grade attainment
	Coefficient
Age	0.601
	(0.677)
Age squared	-0.017
	(0.021)
Household head age	0.005
	(0.013)
Household head age squared	-0.000
	(0.000)
Number of children aged 0 and 5 years	-0.095***
	(0.030)
Number of children aged 6 and 14 years	-0.052**
	(0.023)
Number of children aged 15 and 18 years	0.021
	(0.054)
Professional, technical or Managerial	0.841***
	(0.125)
Clerical	0.614***
	(0.163)
Sales	0.119
	(0.086)
Agricultural - self employed	-0.196***
	(0.074)
Household & Domestic	0.480**
	(0.194)
Services	0.171
	(0.124)
Skilled Manual	0.419**
	(0.166)
Unskilled Manual	0.023
	(0.098)
Embu	0.627
	(0.398)
Kalenjin	-0.456
	(0.290)
Kamba	-0.161
	(0.356)
Kikuyu	0.301
	(0.303)
Kisii	0.815**
	(0.339)
Luhya	0.420
	(0.294)
Luo	0.828**
	(0.389)
Masai	-0.613**
	(0.346)
Meru	0.381
	(0.410)
Mijikenda/Swahili	-0.709**
	(0.319)
Somali	-0.142
	(0.367)

Table 3.9: ***Continued***

Dependent Variable: ***Household Wealth (Ordinal)***	
	School Grade attainment
	Coefficient
Taita	-0.293
	(0.345)
Turkana	-0.797**
	(0.376)
Kuria	0.182
	(0.458)
Protestant & other Christianity	0.103*
	(0.060)
Muslim	0.199
	(0.168)
Head education	0.367***
	(0.106)
Urban	1.543***
	(0.108)
Central	-1.470***
	(0.266)
Coast	-1.361***
	(0.291)
Eastern	-1.636***
	(0.330)
Nyanza	-2.618***
	(0.310)
Rift Valley	-1.522***
	(0.265)
Western	-2.250***
	(0.271)
North Eastern	-2.950***
	(0.338)
Violence	-0.147*
	(0.078)
Polygamy	-0.080
	(0.104)
Constant	0.308
	(0.313)
τ_{1W}	2.836
	(5.596)
τ_{2W}	3.635
	(5.596)
τ_{3W}	4.384
	(5.596)
τ_{4W}	5.493
	(5.596)

*significant at 10%; **significant at 5%; ***significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)
Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 3.10. **Regression Estimates for Rate of Progression during Primary and Secondary School using Age-Qualification.**

Dependent Variable: ***Rate of Progression***	**A: Overall** Number of observations = 8,099 F(39, 3,729) = 61.14	**B: Rural** Number of observations = 6,279 F(37, 2,791) = 51.72	**C: Urban** Number of observations = 1,820 F(38, 937) = 14.68	**D: Female** Number of observations = 3,942 F(32, 2,576) = 36.70	**E: Male** Number of observations = 4,157 F(32, 2,672) = 41.40
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
Age	-0.124***	-0.127***	-0.112***	-0.116***	-0.130***
	(0.006)	(0.006)	(0.010)	(0.008)	(0.008)
Age squared	0.005***	0.005***	0.004***	0.004***	0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	-0.037***	-0.051***	-0.027		
	(0.012)	(0.014)	(0.024)		
Male pre-primary aged siblings	-0.026***	-0.025***	-0.042***	-0.029***	-0.028***
	(0.006)	(0.006)	(0.012)	(0.006)	(0.006)
Male x Male pre-primary aged siblings	-0.004	-0.001	0.008		
	(0.006)	(0.007)	(0.014)		
Male primary school aged siblings	-0.005	-0.002	-0.021**	-0.005	0.001
	(0.004)	(0.005)	(0.010)	(0.005)	(0.005)
Male x Male primary school aged siblings	0.008	0.003	0.032***		
	(0.005)	(0.006)	(0.011)		
Male secondary school aged siblings	-0.004	0.001	-0.038***	-0.005	-0.008
	(0.007)	(0.007)	(0.014)	(0.007)	(0.007)
Male x Male secondary school aged siblings	-0.001	-0.007	0.038**		
	(0.008)	(0.008)	(0.017)		
Female pre-primary aged siblings	-0.010*	-0.014**	0.004	-0.012**	-0.003
	(0.006)	(0.006)	(0.011)	(0.006)	(0.006)
Male x Female pre-primary aged siblings	0.007	0.013*	-0.005		
	(0.007)	(0.007)	(0.014)		
Female primary school aged siblings	0.006	0.005	0.008	0.006	-0.000
	(0.005)	(0.005)	(0.009)	(0.005)	(0.005)
Male x Female primary school aged siblings	-0.004	-0.001	-0.009		
	(0.005)	(0.006)	(0.011)		
Female secondary school aged siblings	-0.004	-0.009	0.008	-0.001	0.000
	(0.006)	(0.008)	(0.008)	(0.007)	(0.009)
Male x Female secondary school aged siblings	0.001	0.009	-0.020		
	(0.008)	(0.009)	(0.016)		

Table 3.10: ***Continued***

Dependent Variable: ***Rate of Progression***	**A: Overall**	**B: Rural**	**C: Urban**	**D: Female**	**E: Male**
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
Deceased mother	-0.022	-0.014	-0.051	-0.010	-0.040
	(0.021)	(0.025)	(0.034)	(0.029)	(0.030)
Deceased father	-0.028*	-0.042***	0.020	-0.031*	-0.011
	(0.014)	(0.015)	(0.023)	(0.018)	(0.018)
Both parents deceased	0.026	0.030	0.014	0.010	0.027
	(0.030)	(0.035)	(0.049)	(0.039)	(0.044)
Protestant & other Christianity	0.005	0.006	0.010	0.011	0.001
	(0.007)	(0.007)	(0.013)	(0.008)	(0.009)
Muslim	-0.036**	-0.003	-0.082***	-0.029	-0.037*
	(0.015)	(0.018)	(0.023)	(0.020)	(0.020)
Poorer household predicted probability	-0.224	0.011	-1.794	-0.850**	0.190
	(0.237)	(0.334)	(4.813)	(0.373)	(0.300)
Middle household predicted probability	0.177	-0.009	0.540	0.631**	-0.225
	(0.202)	(0.354)	(0.843)	(0.282)	(0.244)
Richer household predicted probability	0.021	0.069	-1.191	-0.273	0.092
	(0.153)	(0.256)	(2.242)	(0.202)	(0.190)
Richest household predicted probability	0.153	0.288*	-0.534	0.097	0.004
	(0.146)	(0.149)	(2.001)	(0.164)	(0.154)
Head education	0.181***	0.172***	0.217***	0.162***	0.200***
	(0.014)	(0.015)	(0.025)	(0.021)	(0.017)
Urban	-0.031*			-0.041*	-0.020
	(0.018)			(0.024)	(0.023)
Central	0.050**	0.044	0.043	0.086***	0.017
	(0.023)	(0.035)	(0.035)	(0.030)	(0.030)
Coast	-0.025	-0.056*	0.074**	0.029	-0.065**
	(0.024)	(0.032)	(0.031)	(0.030)	(0.031)
Eastern	-0.011	-0.020	0.104***	0.032	-0.045
	(0.023)	(0.035)	(0.039)	(0.031)	(0.030)
Nyanza	0.018	0.016	0.083**	0.058*	-0.012
	(0.024)	(0.035)	(0.038)	(0.031)	(0.031)
Rift Valley	0.006	0.003	0.017	0.037	-0.017
	(0.022)	(0.035)	(0.032)	(0.028)	(0.029)
Western	-0.003	-0.007	0.047	0.047	-0.046
	(0.023)	(0.035)	(0.039)	(0.030)	(0.031)
North Eastern	0.036		-0.002	0.024	0.033
	(0.031)		(0.063)	(0.045)	(0.039)
Violence	-0.024***	-0.020**	-0.046**	-0.034***	-0.009
	(0.008)	(0.009)	(0.018)	(0.011)	(0.011)

Table 3.10: ***Continued***

Dependent Variable: ***Rate of Progression***	**A: Overall**	**B: Rural**	**C: Urban**	**D: Female**	**E: Male**
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
Polygamy	-0.015	-0.011	-0.029	-0.014	-0.017
	(0.011)	(0.012)	(0.025)	(0.015)	(0.013)
Female circumcision	-0.004	-0.002	-0.008	0.004	-0.006
	(0.008)	(0.008)	(0.017)	(0.009)	(0.011)
Constant	1.218***	1.193***	1.890	1.285***	1.071***
	(0.114)	(0.088)	(1.985)	(0.149)	(0.113)

* significant at 10%; ** significant at 5%; *** significant at 1%

Robust standard errors are in parentheses (adjusted for clustering on household level)

Model Set III: A case of Endogenous Wealth Variable (Continuous). HIV/AIDS variable not included

Table 3.11. Instrumental Variable Probit Estimates for School Enrolment using Age-Qualification.

	3.11a) Dependent Variable: *Primary School Attendance*		3.11b) Dependent Variable: *Secondary School Attendance*	
	Number of observations = 8,209		Number of observations = 2,867	
	Wald χ^2 = 1,075.48		Wald χ^2 = 664.91	
	Prob > χ^2 = 0.000		Prob > χ^2 = 0.000	
	Degrees of freedom = 35		Degrees of freedom = 32	
	Coefficient	**Marginal effect**	**Coefficient**	**Marginal effect**
Age	0.134*	0.013	0.635	0.245
	(0.081)	(0.008)	(0.947)	(0.366)
Age squared	-0.015***	-0.002***	-0.032	-0.012
	(0.004)	(0.000)	(0.029)	(0.011)
Male	0.188	0.019	0.965***	0.356***
	(0.123)	(0.012)	(0.195)	(0.057)
Male pre-primary aged siblings	0.052	0.005	-0.196***	-0.076***
	(0.047)	(0.005)	(0.060)	(0.022)
Male x Male pre-primary aged siblings	-0.082	-0.008	0.196**	0.076**
	(0.053)	(0.005)	(0.085)	(0.033)
Male primary school aged siblings	-0.054*	-0.005*	0.163***	0.063***
	(0.032)	(0.003)	(0.045)	(0.016)
Male x Male primary school aged siblings	0.058	0.006	-0.137**	-0.053**
	(0.048)	(0.005)	(0.062)	(0.022)
Male secondary school aged siblings	0.132**	0.013**	0.120	0.047
	(0.057)	(0.006)	(0.075)	(0.029)
Male x Male secondary school aged siblings	-0.054	-0.005	-0.143	-0.055
	(0.081)	(0.008)	(0.120)	(0.042)
Female pre-primary aged siblings	0.057	0.006	-0.145**	-0.056**
	(0.047)	(0.005)	(0.061)	(0.022)
Male x Female pre-primary aged siblings	-0.012	-0.001)	0.254***	0.098***
	(0.059)	(0.006)	(0.091)	(0.034)
Female primary school aged siblings	-0.006	-0.001	0.162***	0.063***
	(0.035)	(0.003)	(0.051)	(0.017)
Male x Female primary school aged siblings	-0.023	-0.002	-0.042	-0.016
	(0.046)	(0.005)	(0.070)	(0.024)
Female secondary school aged siblings	0.077	0.008	0.287***	0.111***
	(0.085)	(0.009)	(0.089)	(0.028)
Male x Female secondary school aged siblings	-0.000	-0.000	-0.535***	-0.206***
	(0.098)	(0.010)	(0.127)	(0.041)
Deceased mother	-0.033	-0.003		
	(0.173)	(0.018)		
Deceased father	0.201*	0.018*		
	(0.104)	(0.010)		
Both parents deceased	-0.687***	-0.113***		
	(0.242)	(0.059)		
Protestant & other Christianity	0.330***	0.034***	-0.020	-0.008
	(0.069)	(0.008)	(0.068)	(0.025)
Muslim	-0.179	-0.020	-0.123	-0.048
	(0.123)	(0.016)	(0.163)	(0.056)

Table 3.11: ***Continued***

	3.11a) Dependent Variable: *Primary School Attendance*		3.11b) Dependent Variable: *Secondary School Attendance*	
	Coefficient	**Marginal effect**	**Coefficient**	**Marginal effect**
Wealth continuous	0.465***	0.046***	-0.038	-0.016
	(0.079)	(0.016)	(0.108)	(0.025)
Head education	2.020***	0.402***	2.881***	0.700***
	(0.191)	(0.022)	(0.230)	(0.012)
Urban	-0.425***	-0.052**	-0.310**	-0.121**
	(0.143)	(0.028)	(0.201)	(0.053)
Central	1.384***	0.068***	-0.066	-0.026
	(0.241)	(0.018)	(0.223)	(0.067)
Coast	1.453***	0.065***	-0.127	-0.050
	(0.266)	(0.018)	(0.273)	(0.080)
Eastern	1.599***	0.071***	0.033	0.013
	(0.271)	(0.019)	(0.275)	(0.076)
Nyanza	1.838***	0.075***	0.197	0.075
	(0.256)	(0.019)	(0.2840)	(0.074)
Rift Valley	1.245***	0.073***	0.025	0.010
	(0.263)	(0.022)	(0.268)	(0.074)
Western	1.641***	0.070***	0.363*	0.134*
	(0.259)	(0.018)	(0.291)	(0.072)
North Eastern	1.184***	0.057***	0.198	0.074
	(0.380)	(0.019)	(0.430)	(0.113)
Violence	0.050	0.005	-0.161*	-0.063*
	(0.065)	(0.006)	(0.097)	(0.037)
Polygamy	-0.167**	-0.018**	-0.081	-0.032
	(0.076)	(0.009)	(0.122)	(0.048)
Female circumcision	-0.084	-0.008	-0.045	-0.017
	(0.067)	(0.007)	(0.080)	(0.028)
Constant	-1.038**		-4.640	
	(0.495)		(7.797)	
$\ln\sigma$	0.205		0.285	
	(0.021)		(0.013)	
atanh ρ	-0.545		0.002	
	(0.137)		(0.010)	
σ	1.227		1.329	
	(0.026)		(0.018)	
ρ	-0.497		0.002	
	(0.103)		(0.096)	
	Wald test of exogeneity (atanh ρ = 0):		Wald test of exogeneity (atanh ρ = 0):	
	$\chi^2_1 = 15.890$	$\Pr > \chi^2 = 0.000$	$\chi^2_1 = 0.000$	$\Pr > \chi^2 = 0.985$

*significant at 10%; **significant at 5%; ***significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)
Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 3.12. Ordered Probit Estimates for School Attainment using Age-Qualification.[38]

Number of obs = 2,867	Prob > χ^2 = 0.0000	**Marginal effects after Ordered Probit**		
Wald χ^2 = 344.21	Degrees of freedom = 33	**No schooling**	**Some Primary**	**Complete Primary**
Dependent Variable: *School Grade attainment*	**Coefficient**	**Marginal effect**	**Marginal Effect**	**Marginal Effect**
Age	-0.704	0.253	-0.140	-0.113
	(0.738)	(0.265)	(0.147)	(0.119)
Age squared	0.020	-0.007	0.004	0.003
	(0.023)	(0.008)	(0.004)	(0.004)
Male	-0.426***	0.153***	-0.086***	-0.067***
	(0.136)	(0.049)	(0.028)	(0.021)
Male pre-primary aged siblings	0.078	-0.028	0.015	0.012
	(0.054)	(0.019)	(0.011)	(0.009)
Male x Male pre-primary aged siblings	-0.049	0.017	-0.010	-0.008
	(0.067)	(0.024)	(0.013)	(0.011)
Male primary school aged siblings	-0.075**	0.027**	-0.015**	-0.012**
	(0.033)	(0.012)	(0.007)	(0.005)
Male x Male primary school aged siblings	0.109**	-0.039**	0.022**	0.017**
	(0.043)	(0.015)	(0.009)	(0.007)
Male secondary school aged siblings	-0.072	0.026	-0.014	-0.012
	(0.057)	(0.020)	(0.011)	(0.009)
Male x Male secondary school aged siblings	-0.021	0.008	-0.004	-0.003
	(0.079)	(0.028)	(0.016)	(0.013)
Female pre-primary aged siblings	0.082*	-0.029*	0.016*	0.013*
	(0.046)	(0.016)	(0.009)	(0.007)
Male x Female pre-primary aged siblings	-0.076	0.027	-0.015	-0.012
	(0.062)	(0.022)	(0.012)	(0.010)
Female primary school aged siblings	-0.050	0.018	-0.010	-0.008
	(0.038)	(0.014)	(0.008)	(0.006)
Male x Female primary school aged siblings	0.068	-0.025	0.014	0.011
	(0.050)	(0.018)	(0.010)	(0.008)
Female secondary school aged siblings	-0.172***	0.062***	-0.034***	-0.027***
	(0.061)	(0.022)	(0.012)	(0.010)
Male x Female secondary school aged siblings	0.272***	-0.098***	0.054***	0.044***
	(0.089)	(0.032)	(0.018)	(0.014)
Protestant & other Christianity	-0.017	0.006	-0.003	-0.003
	(0.053)	(0.019)	(0.010)	(0.009)
Muslim	0.052	-0.019	0.010	0.009
	(0.129)	(0.046)	(0.024)	(0.022)
Wealth Continuous	-0.119**	0.043**	-0.024**	-0.019**
	(0.052)	(0.019)	(0.010)	(0.008)
Wealth residuals (rivers & vuong)	0.017	-0.006	0.003	0.003
	(0.054)	(0.019)	(0.011)	(0.009)
Head education	3.411***	-0.787***	0.628***	0.160***
	(0.281)	(0.009)	(0.011)	(0.008)
Urban	0.171	-0.060	0.031*	0.029
	(0.109)	(0.037)	(0.018)	(0.020)
Central	-0.345**	0.129**	-0.082*	-0.047**
	(0.167)	(0.065)	(0.046)	(0.019)
Coast	-0.152	0.056	-0.033	-0.022
	(0.189)	(0.071)	(0.045)	(0.026)
Eastern	-0.134	0.049	-0.029	-0.020
	(0.179)	(0.067)	(0.042)	(0.025)

[38] Used Rivers Vuong residuals for continuous wealth variable.

Table 3.12: ***Continued***

		Marginal effects after Ordered Probit		
		No schooling	**Some Primary**	**Complete Primary**
Dependent Variable: ***School Grade attainment***	**Coefficient**	**Marginal effect**	**Marginal Effect**	**Marginal Effect**
Nyanza	-0.556***	0.212***	-0.143***	-0.068***
	(0.181)	(0.071)	(0.055)	(0.017)
Rift Valley	-0.163	0.060	-0.036	-0.024
	(0.178)	(0.067)	(0.042)	(0.025)
Western	-0.362**	0.136**	-0.088**	-0.048**
	(0.181)	(0.070)	(0.051)	(0.020)
North Eastern	-0.659**	0.254**	-0.183**	-0.071***
	(0.262)	(0.103)	(0.085)	(0.018)
Violence	0.091	-0.032	0.017	0.015
	(0.069)	(0.024)	(0.012)	(0.012)
Polygamy	-0.022	0.008	-0.004	-0.003
	(0.079)	(0.028)	(0.016)	(0.012)
Female circumcision	0.043	-0.015	0.008	0.007
	(0.056)	(0.020)	(0.011)	(0.009)
τ_{1E}	-4.078			
	(6.044)			
τ_{2E}	-2.268			
	(6.044)			

*significant at 10%; **significant at 5%; ***significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)
Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 3.13. **IVRegression Estimates for Rate of Progression during Primary and Secondary School using Age-Qualification.**[39]

Dependent Variable: ***Rate of Progression***	**A: Overall** Number of observations = 8,030 F(35, 3,701) = 60.51	**B: Rural** Number of observations = 6,225 F(33, 2,767) = 45.37	**C: Urban** Number of observations = 1,805 F(34, 933) = 15.54	**D: Female** Number of observations = 3,911 F(28, 2,557) = 37.33	**E: Male** Number of observations = 4,119 F(28, 2,649) = 42.41
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
Age	-0.122***	-0.124***	-0.111***	-0.116***	-0.126***
	(0.005)	(0.006)	(0.011)	(0.007)	(0.008)
Age squared	0.005***	0.005***	0.004***	0.005***	0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	-0.041***	-0.050***	-0.034		
	(0.012)	(0.013)	(0.022)		
Male pre-primary aged siblings	-0.023***	-0.024***	-0.029**	-0.027***	-0.026***
	(0.005)	(0.006)	(0.012)	(0.005)	(0.006)
Male x Male pre-primary aged siblings	-0.005	-0.004	-0.002		
	(0.006)	(0.007)	(0.014)		
Male primary school aged siblings	-0.008*	-0.005	-0.023**	-0.008*	0.002
	(0.004)	(0.005)	(0.010)	(0.004)	(0.005)
Male x Male primary school aged siblings	0.009*	0.006	0.036***		
	(0.005)	(0.006)	(0.011)		
Male secondary school aged siblings	-0.006	-0.004	-0.033**	-0.008	-0.008
	(0.006)	(0.007)	(0.014)	(0.006)	(0.007)
Male x Male secondary school aged siblings	-0.005	-0.005	0.026		
	(0.008)	(0.008)	(0.017)		
Female pre-primary aged siblings	-0.008	-0.011*	0.008	-0.010*	-0.001
	(0.005)	(0.006)	(0.010)	(0.005)	(0.006)
Male x Female pre-primary aged siblings	0.005	0.009	-0.006		
	(0.006)	(0.007)	(0.013)		
Female primary school aged siblings	0.003	0.001	0.004	0.003	0.001
	(0.004)	(0.004)	(0.009)	(0.004)	(0.004)
Male x Female primary school aged siblings	-0.002	-0.001	-0.005		
	(0.005)	(0.005)	(0.011)		
Female secondary school aged siblings	-0.006	-0.012	0.004	-0.007	0.001
	(0.006)	(0.008)	(0.009)	(0.006)	(0.009)
Male x Female secondary school aged siblings	0.005	0.006	-0.006		
	(0.008)	(0.009)	(0.015)		

[39] These estimates are obtained using IV procedure for school grade progression.

Table 3.13: ***Continued***

Dependent Variable: ***Rate of Progression***	**A: Overall**	**B: Rural**	**C: Urban**	**D: Female**	**E: Male**
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
Deceased mother	-0.027	-0.014	-0.058*	-0.021	-0.030
	(0.021)	(0.024)	(0.035)	(0.027)	(0.032)
Deceased father	-0.005	-0.015	0.034	-0.009	0.002
	(0.012)	(0.014)	(0.023)	(0.016)	(0.016)
Both parents deceased	0.013	0.004	0.003	0.009	0.011
	(0.031)	(0.035)	(0.053)	(0.041)	(0.045)
Protestant & other Christianity	0.009	0.009	0.008	0.014*	0.004
	(0.007)	(0.008)	(0.013)	(0.008)	(0.009)
Muslim	-0.057***	-0.012	-0.113***	-0.052**	-0.061***
	(0.016)	(0.020)	(0.023)	(0.021)	(0.020)
Wealth continuous	0.063***	0.086***	0.030***	0.055***	0.073***
	(0.006)	(0.008)	(0.007)	(0.007)	(0.009)
Head education	0.187***	0.174***	0.213***	0.168***	0.202***
	(0.013)	(0.014)	(0.025)	(0.020)	(0.017)
Urban	0.005			0.002	0.005
	(0.013)			(0.016)	(0.018)
Central	0.174***	0.026	0.032	0.165***	0.182***
	(0.020)	(0.029)	(0.026)	(0.023)	(0.028)
Coast	0.076***	-0.071***	0.029	0.078***	0.079**
	(0.024)	(0.025)	(0.022)	(0.029)	(0.032)
Eastern	0.110***	-0.029	0.052*	0.102***	0.117***
	(0.022)	(0.028)	(0.031)	(0.027)	(0.031)
Nyanza	0.118***	-0.018	0.026	0.105***	0.132***
	(0.022)	(0.029)	(0.024)	(0.026)	(0.031)
Rift Valley	0.103***	-0.042	0.004	0.080***	0.126***
	(0.021)	(0.029)	(0.027)	(0.025)	(0.030)
Western	0.094***	-0.045	0.004	0.088***	0.101***
	(0.022)	(0.028)	(0.027)	(0.027)	(0.031)
North Eastern	0.171***		0.065	0.120***	0.211***
	(0.034)		(0.042)	(0.045)	(0.044)
Violence	-0.015*	-0.017*	-0.023	-0.027**	-0.003
	(0.008)	(0.009)	(0.018)	(0.011)	(0.010)
Polygamy	-0.003	-0.007	-0.009	-0.002	-0.006
	(0.011)	(0.012)	(0.025)	(0.015)	(0.013)
Female circumcision	0.002	0.006	-0.001	0.008	-0.003
	(0.007)	(0.008)	(0.016)	(0.009)	(0.009)
constant	1.177***	1.371***	1.209***	1.174***	1.137***
	(0.039)	(0.046)	(0.066)	(0.049)	(0.057)

* significant at 10%; ** significant at 5%; *** significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)

Model Set IV: A case of Endogenous Wealth Variable with HIV/AIDS variable included[40]

Table 3.14. Estimates for School Enrolment using Age-Qualification.

	3.14a) Dependent Variable: *Primary School Attendance*		3.14b) Dependent Variable: *Secondary School Attendance*	
	Number of observations = 3,604		Number of observation = 1,219	
	Wald χ^2 = 600.43		Wald χ^2 = 327.61	
	Prob > χ^2 = 0.000		Prob > χ^2 = 0.0000	
	Degrees of freedom = 36		Degrees of freedom = 33	
	Coefficient	**Marginal Effect**	**Coefficient**	**Marginal Effect**
Age	0.145	0.011	0.852	0.321
	(0.125)	(0.010)	(1.431)	(0.539)
Age squared	-0.016**	-0.001**	-0.040	-0.015
	(0.007)	(0.001)	(0.043)	(0.016)
Male	0.069	0.005	1.045***	0.369***
	(0.214)	(0.016)	(0.273)	(0.086)
Male pre-primary aged siblings	0.065	0.005	-0.299***	-0.113***
	(0.068)	(0.006)	(0.093)	(0.035)
Male x Male pre-primary aged siblings	-0.156**	-0.012*	0.280**	0.105**
	(0.078)	(0.007)	(0.135)	(0.051)
Male primary school aged siblings	-0.084	-0.006*	0.221***	0.083***
	(0.054)	(0.004)	(0.080)	(0.030)
Male x Male primary school aged siblings	0.064	0.005	-0.179*	-0.067*
	(0.074)	(0.006)	(0.107)	(0.040)
Male secondary school aged siblings	0.237**	0.018*	0.057	0.022
	(0.108)	(0.009)	(0.120)	(0.045)
Male x Male secondary school aged siblings	0.096	0.007	-0.184	-0.069
	(0.157)	(0.012)	(0.192)	(0.072)
Female pre-primary aged siblings	-0.081	-0.006	-0.317***	-0.119***
	(0.086)	(0.006)	(0.082)	(0.031)
Male x Female pre-primary aged siblings	0.196*	0.015*	0.287**	0.108**
	(0.106)	(0.008)	(0.142)	(0.054)
Female primary school aged siblings	-0.039	-0.003	0.187**	0.070**
	(0.057)	(0.004)	(0.075)	(0.028)
Male x Female primary school aged siblings	-0.017	-0.001	-0.100	-0.038
	(0.071)	(0.005)	(0.105)	(0.040)
Female secondary school aged siblings	0.123	0.009	0.181*	0.068*
	(0.142)	(0.011)	(0.103)	(0.039)
Male x Female secondary school aged siblings	0.050	0.004	-0.654***	-0.246***
	(0.159)	(0.012)	(0.153)	(0.057)
Deceased mother	-0.202	-0.018		
	(0.232)	(0.025)		
Deceased father	0.211	0.014		
	(0.166)	(0.011)		
Both parents deceased	-0.129	-0.011		
	(0.380)	(0.036)		

[40] The estimates are obtained as in Model set III, but with additional variable for any HIV/AIDS individual in the household (i.e with missing observations).

Table 3.14: ***Continued***

	3.14a) Dependent Variable: *Primary School Attendance*		3.14b) Dependent Variable: *Secondary School Attendance*	
	Coefficient	**Marginal Effect**	**Coefficient**	**Marginal Effect**
Protestant & other Christianity	0.354***	0.029***	-0.035	-0.013
	(0.101)	(0.010)	(0.103)	(0.039)
Muslim	0.176	0.012	0.046	0.017
	(0.169)	(0.011)	(0.241)	(0.090)
Wealth continuous	0.601***	0.046**	-0.215	-0.081
	(0.099)	(0.018)	(0.160)	(0.061)
Head education	2.052***	0.365***	2.629***	0.695***
	(0.233)	(0.028)	(0.385)	(0.028)
Urban	-0.534***	-0.055	-0.050	-0.019
	(0.184)	(0.034)	(0.296)	(0.112)
Central	1.540***	0.055***	-0.295	-0.114
	(0.392)	(0.020)	(0.316)	(0.124)
Coast	1.501***	0.048***	-0.623*	-0.244*
	(0.400)	(0.018)	(0.374)	(0.145)
Eastern	2.246***	0.064***	-0.325	-0.126
	(0.413)	(0.022)	(0.397)	(0.157)
Nyanza	2.444***	0.071***	-0.163	-0.062
	(0.390)	(0.023)	(0.406)	(0.158)
Rift Valley	1.591***	0.065	-0.221	-0.085
	(0.417)	(0.026)	(0.381)	(0.149)
Western	1.723***	0.055	0.155	0.057
	(0.426)	(0.020)	(0.439)	(0.158)
North Eastern	1.319**	0.044	-0.434	-0.170
	(0.526)	(0.018)	(0.641)	(0.254)
Violence	0.122	0.009	-0.131	-0.050
	(0.093)	(0.007)	(0.148)	(0.057)
Any HIV patient	-0.197	-0.017	-0.417***	-0.162***
	(0.175)	(0.017)	(0.128)	(0.050)
Polygamy	-0.264**	-0.024*	-0.085	-0.032
	(0.111)	(0.013)	(0.174)	(0.067)
Female circumcision	-0.204*	-0.016**	-0.136	-0.052
	(0.107)	(0.008)	(0.126)	(0.048)
constant	-1.241*		-5.283	
	(0.734)		(11.799)	
$\ln\sigma$	0.130***		0.275***	
	(0.032)		(0.038)	
atanh ρ	-0.560***		0.242	
	(0.162)		(0.255)	
σ	1.138		1.317	
	(0.036)		(0.050)	
ρ	-0.508		0.237	
	(0.120)		(0.240)	
	Wald test of exogeneity (atanh ρ = 0):		Wald test of exogeneity (atanh ρ = 0):	
	$\chi^2_1 = 12.010$ Pr > $\chi^2 = 0.000$		$\chi^2_1 = 0.900$ Pr > $\chi^2 = 0.324$	

*significant at 10%; **significant at 5%; ***significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)
Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 3.15. Ordered Probit Estimates for School Attainment using Age-Qualification

Number of obs = 1,219	Prob > χ^2 = 0.0000	**Marginal effects after ordered probit**		
Wald χ^2 = 162.89	Degrees of freedom = 34	**No schooling**	**Some Primary**	**Complete Primary**
Dependent Variable: *School Grade attainment*	**Coefficient**	**Marginal effect**	**Marginal Effect**	**Marginal Effect**
Age	-1.415	0.484	-0.237	-0.247
	(1.134)	(0.387)	(0.190)	(0.200)
Age squared	0.042	-0.014	0.007	0.007
	(0.035)	(0.012)	(0.006)	(0.006)
Male	-0.044	0.015	-0.007	-0.008
	(0.189)	(0.065)	(0.032)	(0.033)
Male pre-primary aged siblings	0.112	-0.038	0.019	0.020
	(0.078)	(0.027)	(0.013)	(0.014)
Male x Male pre-primary aged siblings	-0.105	0.036	-0.018	-0.018
	(0.102)	(0.035)	(0.017)	(0.018)
Male primary school aged siblings	-0.061	0.021	-0.010	-0.011
	(0.050)	(0.017)	(0.008)	(0.009)
Male x Male primary school aged siblings	0.074	-0.025	0.012	0.013
	(0.065)	(0.022)	(0.011)	(0.011)
Male secondary school aged siblings	0.015	-0.005	0.002	0.003
	(0.088)	(0.030)	(0.015)	(0.015)
Male x Male secondary school aged siblings	-0.140	0.048	-0.023	-0.024
	(0.112)	(0.038)	(0.019)	(0.019)
Female pre-primary aged siblings	0.156**	-0.053**	0.026**	0.027**
	(0.066)	(0.022)	(0.011)	(0.012)
Male x Female pre-primary aged siblings	-0.156*	0.053*	-0.026*	-0.027*
	(0.094)	(0.032)	(0.016)	(0.017)
Female primary school aged siblings	-0.096*	0.033*	-0.016*	-0.017*
	(0.056)	(0.019)	(0.010)	(0.010)
Male x Female primary school aged siblings	0.126*	-0.043*	0.021	0.022*
	(0.073)	(0.025)	(0.013)	(0.013)
Female secondary school aged siblings	-0.029	0.010	-0.005	-0.005
	(0.081)	(0.028)	(0.013)	(0.014)
Male x Female secondary school aged siblings	-0.011	0.004	-0.002	-0.002
	(0.127)	(0.043)	(0.021)	(0.022)
Protestant & other Christianity	-0.093	0.031	-0.015	-0.016
	(0.078)	(0.026)	(0.012)	(0.014)
Muslim	-0.518**	0.193**	-0.124**	-0.068***
	(0.204)	(0.080)	(0.061)	(0.021)
Wealth continuous	0.037	-0.013	0.006	0.007
	(0.079)	(0.027)	(0.013)	(0.014)
Wealth residuals (Rivers & Vuong)	-0.139*	0.048*	-0.023	-0.024*
	(0.082)	(0.028)	(0.014)	(0.014)
Head education	3.015***	-0.779***	0.626***	0.152***
	(0.427)	(0.021)	(0.023)	(0.011)
Urban	0.053	-0.018	0.009	0.009
	(0.159)	(0.053)	(0.025)	(0.029)
Central	0.036	-0.012	0.006	0.006
	(0.258)	(0.087)	(0.041)	(0.046)
Coast	0.457	-0.138	0.039	0.010
	(0.295)	(0.077)	(0.010)	(0.077)
Eastern	0.225	-0.073	0.029	0.044
	(0.284)	(0.087)	(0.027)	(0.061)
Nyanza	-0.111	0.039	-0.020	-0.018
	(0.286)	(0.102)	(0.056)	(0.045)

Table 3.15: ***Continued***

		Marginal effects after ordered probit		
		No schooling	**Some Primary**	**Complete Primary**
Dependent Variable: ***School Grade attainment***	**Coefficient**	**Marginal effect**	**Marginal Effect**	**Marginal Effect**
Rift Valley	0.206	-0.068	0.029	0.039
	(0.274)	(0.086)	(0.031)	(0.056)
Western	0.062	-0.021	0.010	0.011
	(0.285)	(0.095)	(0.043)	(0.053)
North Eastern	0.160	-0.052	0.022	0.031
	(0.440)	(0.137)	(0.046)	(0.092)
Violence	0.135	-0.045	0.020	0.025
	(0.105)	(0.034)	(0.014)	(0.021)
Any HIV patient	0.101	-0.034	0.015	0.018
	(0.115)	(0.038)	(0.016)	(0.022)
Polygamy	-0.006	0.002	-0.001	-0.001
	(0.117)	(0.040)	.	(0.020)
Female circumcision	0.004	-0.001	0.001	0.001
	(0.084)	(0.029)	(0.014)	(0.015)
τ_{1E}	-9.452			
	(9.279)			
τ_{2E}	-7.610			
	(9.281)			

*significant at 10%; **significant at 5%; ***significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)
Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 3.16. IV Regression Estimates for Rate of Progression during Primary and Secondary School using Age-Qualification

Dependent Variable: ***Rate of Progression***	**A: Overall** Number of observations = 3,541 F(36, 1,628) = 27.95	**B: Rural** Number of observations = 2,798 F(34, 1,256) = 19.86	**C: Urban** Number of observations = 743 F(35, 371) = 7.84	**D: Female** Number of observations = 1,744 F(29, 1,136) = 14.79	**E: Male** Number of observations = 1,797 F(29, 1,162) = 23.34
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
Age	-0.118***	-0.112***	-0.133***	-0.110***	-0.123***
	(0.008)	(0.009)	(0.016)	(0.011)	(0.011)
Age squared	0.005***	0.004***	0.005***	0.004***	0.005***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Male	-0.028	-0.041**	0.019		
	(0.017)	(0.020)	(0.031)		
Male pre-primary aged siblings	-0.030***	-0.030***	-0.023	-0.034***	-0.041***
	(0.008)	(0.009)	(0.017)	(0.008)	(0.008)
Male x Male pre-primary aged siblings	-0.013	-0.015	-0.009		
	(0.009)	(0.011)	(0.020)		
Male primary school aged siblings	-0.008	-0.005	-0.022	-0.008	0.001
	(0.006)	(0.007)	(0.014)	(0.006)	(0.006)
Male x Male primary school aged siblings	0.008	0.006	0.019		
	(0.008)	(0.009)	(0.017)		
Male secondary school aged siblings	-0.010	-0.017	-0.016	-0.011	-0.010
	(0.010)	(0.011)	(0.020)	(0.009)	(0.010)
Male x Male secondary school aged siblings	0.002	0.012	-0.003		
	(0.011)	(0.013)	(0.021)		
Female pre-primary aged siblings	-0.015*	-0.017*	0.004	-0.018**	-0.007
	(0.008)	(0.009)	(0.019)	(0.009)	(0.008)
Male x Female pre-primary aged siblings	0.006	0.010	-0.011		
	(0.009)	(0.011)	(0.021)		
Female primary school aged siblings	0.009	0.006	0.015	0.009	-0.001
	(0.006)	(0.007)	(0.011)	(0.006)	(0.006)
Male x Female primary school aged siblings	-0.010	-0.009	-0.014		
	(0.007)	(0.008)	(0.013)		
Female secondary school aged siblings	-0.008	-0.013	0.008	-0.007	-0.015
	(0.009)	(0.012)	(0.012)	(0.010)	(0.012)
Male x Female secondary school aged siblings	-0.010	-0.008	-0.016		
	(0.011)	(0.013)	(0.019)		
Deceased mother	-0.053*	-0.074**	-0.027	-0.048	-0.064
	(0.029)	(0.038)	(0.041)	(0.038)	(0.043)

Table 3.16: ***Continued***

Dependent Variable: *Rate of Progression*	A: Overall	B: Rural	C: Urban	D: Female	E: Male
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
Deceased father	-0.008	-0.010	0.049	-0.002	-0.012
	(0.018)	(0.020)	(0.036)	(0.024)	(0.021)
Both parents deceased	0.017	0.012	-0.025	0.005	0.035
	(0.043)	(0.051)	(0.075)	(0.056)	(0.061)
Protestant & other Christianity	0.014	0.012	0.021	0.013	0.013
	(0.010)	(0.011)	(0.020)	(0.013)	(0.012)
Muslim	-0.056**	-0.018	-0.119***	-0.057*	-0.053*
	(0.024)	(0.029)	(0.038)	(0.033)	(0.029)
Wealth continuous	0.047***	0.079***	0.021**	0.049***	0.047***
	(0.008)	(0.013)	(0.009)	(0.010)	(0.010)
Head education	0.182***	0.173***	0.205***	0.149***	0.200***
	(0.020)	(0.023)	(0.041)	(0.034)	(0.024)
Urban	0.045***			0.021	0.063***
	(0.017)			(0.022)	(0.020)
Central	0.124***	0.011	0.010	0.131***	0.112***
	(0.025)	(0.045)	(0.036)	(0.032)	(0.031)
Coast	0.034	-0.083**	0.020	0.071*	-0.001
	(0.030)	(0.041)	(0.030)	(0.040)	(0.036)
Eastern	0.063**	-0.036	0.022	0.078**	0.041
	(0.030)	(0.043)	(0.041)	(0.037)	(0.038)
Nyanza	0.071**	-0.023	-0.003	0.098***	0.039
	(0.029)	(0.043)	(0.032)	(0.036)	(0.036)
Rift Valley	0.064**	-0.041	-0.016	0.074**	0.046
	(0.028)	(0.044)	(0.036)	(0.035)	(0.034)
Western	0.056*	-0.034	-0.054	0.077**	0.030
	(0.030)	(0.043)	(0.040)	(0.038)	(0.036)
North Eastern	0.109**		0.018	0.135*	0.080
	(0.047)		(0.067)	(0.070)	(0.053)
Violence	-0.014	-0.019	-0.011	-0.027*	-0.004
	(0.012)	(0.013)	(0.027)	(0.016)	(0.014)
Any HIV patient	0.004	0.009	-0.040*	-0.002	0.007
	(0.014)	(0.018)	(0.023)	(0.017)	(0.017)
Polygamy	0.019	0.013	0.067**	0.023	0.016
	(0.015)	(0.017)	(0.034)	(0.022)	(0.018)
Female circumcision	-0.003	-0.001	-0.003	-0.006	-0.002
	(0.011)	(0.012)	(0.027)	(0.014)	(0.014)
Constant	1.200***	1.307***	1.356***	1.187***	1.193***
	(0.055)	(0.070)	(0.091)	(0.071)	(0.078)

* significant at 10%; ** significant at 5%; *** significant at 1%

Robust standard errors are in parentheses (adjusted for clustering on household level)

Model Set V: A case of Endogenous Wealth Variable with Selection model correction for HIV/AIDS variable[41]

Table 3.17. Heckman 2-step Estimates for School Enrolment using Age-Qualification

	3.17a) Dependent Variable: *Primary School Attendance*		3.17b) Dependent Variable: *Secondary School Attendance*	
	Number of observations = 4,009		Number of observations = 1,368	
	Censored observations = 401		Censored observations = 146	
	Uncensored observations = 3,608		Uncensored observations = 1,222	
	Wald χ^2 = 529.16		Wald χ^2 = 243.990	
	Prob > χ^2 = 0.000		Prob > χ^2 = 0.000	
	Degrees of freedom = 35		Degrees of freedom = 32	
	Coefficient	**Marginal effect**	**Coefficient**	**Marginal effect**
Age	0.108	0.005	0.854	0.409
	(0.138)	(0.007)	(1.401)	(0.537)
Age squared	-0.016**	-0.001**	-0.040	-0.018
	(0.007)	(0.000)	(0.042)	(0.016)
Male	0.103	0.005	0.962***	0.378***
	(0.244)	(0.012)	(0.284)	(0.092)
Male pre-primary aged siblings	0.048	0.002	-0.331***	-0.122***
	(0.071)	(0.004)	(0.091)	(0.035)
Male x Male pre-primary aged siblings	-0.139	-0.008*	0.316**	0.117**
	(0.087)	(0.004)	(0.130)	(0.050)
Male primary school aged siblings	-0.152***	-0.008***	0.259***	0.096***
	(0.057)	(0.003)	(0.071)	(0.027)
Male x Male primary school aged siblings	0.091	0.005	-0.223**	-0.083**
	(0.081)	(0.004)	(0.100)	(0.038)
Male secondary school aged siblings	0.262**	0.013*	0.065	0.024
	(0.121)	(0.007)	(0.117)	(0.045)
Male x Male secondary school aged siblings	0.108	0.006	-0.222	-0.083
	(0.178)	(0.009)	(0.193)	(0.074)
Female pre-primary aged siblings	-0.130	-0.006	-0.335***	-0.118***
	(0.086)	(0.004)	(0.081)	(0.031)
Male x Female pre-primary aged siblings	0.235**	0.012*	0.327**	0.115**
	(0.113)	(0.006)	(0.134)	(0.052)
Female primary school aged siblings	-0.048	-0.003	0.151**	0.069**
	(0.062)	(0.003)	(0.075)	(0.028)
Male x Female primary school aged siblings	-0.013	-0.001	-0.112	-0.046
	(0.078)	(0.004)	(0.103)	(0.039)
Female secondary school aged siblings	0.134	0.006	0.132	0.078*
	(0.139)	(0.007)	(0.105)	(0.041)
Male x Female secondary school aged siblings	-0.052	-0.002	-0.499***	-0.221***
	(0.174)	(0.009)	(0.155)	(0.058)
Deceased mother	-0.137	-0.008		
	(0.240)	(0.016)		
Deceased father	-0.001	-0.001		
	(0.164)	(0.008)		
Both parents deceased	0.045	0.002		
	(0.403)	(0.019)		

[41] The estimates are obtained as in Model set IV but used Heckman probit selection model for HIV.

Table 3.17: ***Continued***

	3.17a) Dependent Variable: *Primary School Attendance*		3.17b) Dependent Variable: *Secondary School Attendance*	
	Coefficient	**Marginal effect**	**Coefficient**	**Marginal effect**
Protestant & other Christianity	0.370***	0.019***	-0.090	-0.019
	(0.110)	(0.006)	(0.104)	(0.039)
Muslim	0.054	0.004	0.090	0.028
	(0.196)	(0.009)	(0.252)	(0.093)
Wealth Continuous	0.429***	0.021***	-0.038	-0.020
	(0.100)	(0.006)	(0.087)	(0.033)
Wealth residuals (Rivers and Voung)	-0.216**	-0.011**	-0.007	0.001
	(0.094)	(0.005)	(0.096)	(0.037)
Head education	2.452***	0.405***	2.386***	0.704***
	(0.157)	(0.031)	(0.363)	(0.026)
Urban	-0.427**	-0.026	-0.287	-0.107
	(0.204)	(0.017)	(0.182)	(0.072)
Central	1.032**	0.028***	-0.235	-0.077
	(0.401)	(0.008)	(0.258)	(0.101)
Coast	1.090**	0.026***	-0.454	-0.172
	(0.428)	(0.007)	(0.305)	(0.122)
Eastern	1.708***	0.035***	-0.133	-0.045
	(0.443)	(0.008)	(0.306)	(0.119)
Nyanza	2.028***	0.037***	-0.066	0.005
	(0.443)	(0.008)	(0.300)	(0.112)
Rift Valley	1.077**	0.031***	-0.117	-0.023
	(0.430)	(0.010)	(0.288)	(0.110)
Western	1.303***	0.029***	0.344	0.142
	(0.458)	(0.007)	(0.307)	(0.099)
North Eastern	0.575	0.017	-0.277	-0.091
	(0.526)	(0.011)	(0.539)	(0.214)
Violence	0.202*	0.008*	-0.228	-0.087
	(0.105)	(0.005)	(0.140)	(0.056)
Any HIV patient	-0.183	-0.009	-0.412***	-0.159***
	(0.180)	(0.009)	(0.122)	(0.047)
Polygamy	-0.212*	-0.011	-0.189	-0.060
	(0.121)	(0.008)	(0.163)	(0.064)
Female circumcision	-0.073	-0.004	-0.048	-0.019
	(0.138)	(0.007)	(0.155)	(0.061)
Constant	-0.394		-5.046	
	(0.719)		(11.570)	
atanh ρ	0.268*		0.736*	
	(0.152)		(0.390)	
ρ	0.262		0.626	
	(0.142)		(0.237)	
	Wald test of independent equations (ρ = 0):		Wald test of independent equations(ρ =0):	
	$\chi_1^2 = 3.09$ Pr > $\chi^2 = 0.079$		$\chi_1^2 = 3.56$	Pr > $\chi^2 = 0.059$

Table 3.18. **Heckman 2-step for School Enrolment:** ***Equation for Household member consent.***

Dependent Variable: ***Household member consent***		
	Primary School Attendance	***Secondary School Attendance***
	Coefficient	**Coefficient**
Age	-0.007	1.757
	(0.093)	(1.613)
Age squared	0.000	-0.053
	(0.004)	(0.049)
Male	-0.020	0.672*
	(0.174)	(0.399)
Male pre-primary aged siblings	0.048	0.137
	(0.072)	(0.108)
Male x Male pre-primary aged siblings	0.113	-0.126
	(0.077)	(0.142)
Male primary school aged siblings	0.053	-0.105
	(0.061)	(0.080)
Male x Male primary school aged siblings	-0.115	0.084
	(0.078)	(0.107)
Male secondary school aged siblings	0.067	-0.040
	(0.098)	(0.143)
Male x Male secondary school aged siblings	-0.116	0.068
	(0.108)	(0.251)
Female pre-primary aged siblings	-0.028	0.261*
	(0.085)	(0.140)
Male x Female pre-primary aged siblings	0.063	-0.266*
	(0.082)	(0.152)
Female primary school aged siblings	0.021	0.237**
	(0.074)	(0.104)
Male x Female primary school aged siblings	0.124	-0.059
	(0.089)	(0.119)
Female secondary school aged siblings	0.171*	0.616**
	(0.102)	(0.265)
Male x Female secondary school aged siblings	-0.122	-0.616**
	(0.119)	(0.286)
Deceased mother	0.108	
	(0.203)	
Deceased father	0.180	
	(0.145)	
Protestant & other Christianity	0.189*	0.337**
	(0.103)	(0.134)
Muslim	-0.310	-0.132
	(0.223)	(0.281)
Wealth Continuous	0.054	-0.112
	(0.078)	(0.101)
Wealth residuals	-0.068	0.090
	(0.085)	(0.111)
Head education	0.001	0.340*
	(0.101)	(0.198)
Urban	-0.324*	0.135
	(0.176)	(0.246)
Central	0.311	0.395
	(0.237)	(0.300)
Coast	0.641**	0.251
	(0.281)	(0.343)
Eastern	0.374	0.168
	(0.281)	(0.355)

Table 3.18: ***Continued***

Dependent Variable: ***Household member consent***

	Primary School Attendance	***Secondary School Attendance***
	Coefficient	**Coefficient**
Nyanza	1.154***	1.228***
	(0.341)	(0.384)
Rift Valley	0.589**	0.665*
	(0.279)	(0.376)
Western	0.903***	0.695*
	(0.304)	(0.379)
North Eastern	0.846**	0.552
	(0.393)	(0.530)
Violence	0.375***	0.081
	(0.119)	(0.165)
Polygamy	-0.183	0.370
	(0.138)	(0.265)
Union	0.031	-0.517*
	(0.187)	(0.304)
Working away	-0.302***	-0.396***
	(0.101)	(0.125)
Constant	0.723	-15.029
	(0.508)	(13.278)

*significant at 10%; **significant at 5%; ***significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)
Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 3.19. Ordered Probit Estimates for School Attainment using Age-Qualification[42]

Number of obs = 1,222 Prob > χ^2 = 0.0000		**Marginal effects after Ordered Probit**		
Wald χ^2 = 151.96 Degrees of freedom = 33		**No schooling**	**Some Primary**	**Complete Primary**
Dependent Variable: *School Grade attainment*	**Coefficient**	**Marginal effect**	**Marginal Effect**	**Marginal Effect**
Age	-1.552	0.532	-0.262	-0.271
	(1.135)	(0.389)	(0.191)	(0.201)
Age squared	0.047	-0.016	0.008	0.008
	(0.035)	(0.012)	(0.006)	(0.006)
Male	-0.089	0.030	-0.015	-0.015
	(0.192)	(0.066)	(0.033)	(0.033)
Male pre-primary aged siblings	0.156**	-0.054**	0.026**	0.027**
	(0.077)	(0.026)	(0.013)	(0.014)
Male x Male pre-primary aged siblings	-0.060	0.021	-0.010	-0.010
	(0.105)	(0.036)	(0.018)	(0.018)
Male primary school aged siblings	-0.058	0.020	-0.010	-0.010
	(0.050)	(0.017)	(0.008)	(0.009)
Male x Male primary school aged siblings	0.055	-0.019	0.009	0.010
	(0.066)	(0.023)	(0.011)	(0.012)
Male secondary school aged siblings	-0.004	0.001	-0.001	-0.001
	(0.089)	(0.030)	(0.015)	(0.015)
Male x Male secondary school aged siblings	-0.123	0.042	-0.021	-0.021
	(0.113)	(0.039)	(0.019)	(0.020)
Female pre-primary aged siblings	0.199***	-0.068***	0.034***	0.035***
	(0.067)	(0.023)	(0.012)	(0.012)
Male x Female pre-primary aged siblings	-0.189*	0.065*	-0.032*	-0.033*
	(0.097)	(0.033)	(0.017)	(0.017)
Female primary school aged siblings	-0.066	0.023	-0.011	-0.012
	(0.059)	(0.020)	(0.010)	(0.010)
Male x Female primary school aged siblings	0.160**	-0.055**	0.027**	0.028**
	(0.072)	(0.025)	(0.013)	(0.013)
Female secondary school aged siblings	0.020	-0.007	0.003	0.003
	(0.086)	(0.030)	(0.015)	(0.015)
Male x Female secondary school aged siblings	-0.081	0.028	-0.014	-0.014
	(0.128)	(0.044)	(0.022)	(0.022)
Protestant & other Christianity	-0.005	0.002	-0.001	-0.001
	(0.090)	(0.031)	(0.015)	(0.016)
Muslim	-0.605***	0.227***	-0.151**	-0.076***
	(0.211)	(0.083)	(0.065)	(0.020)
Wealth continuous	-0.028	0.009	-0.005	-0.005
	(0.074)	(0.025)	(0.012)	(0.013)
Wealth residuals (rivers and voung)	-0.074	0.025	-0.012	-0.013
	(0.077)	(0.026)	(0.013)	(0.013)
Head education	3.017***	-0.778***	0.625***	0.153***
	(0.428)	(0.021)	(0.023)	(0.011)
Urban	0.016	-0.006	0.003	0.003
	(0.170)	(0.058)	(0.028)	(0.030)
Central	0.222	-0.073	0.030	0.042
	(0.288)	(0.090)	(0.031)	(0.060)
Coast	0.785**	-0.214***	0.019	0.195*
	(0.363)	(0.073)	(0.045)	(0.115)

[42] We constructed Rivers and Vuong residuals manually and computed IMR for HIV variable for the ordered probit for school attainment.

Table 3.19: ***Continued***

		Marginal effects after Ordered Probit		
		No schooling	**Some Primary**	**Complete Primary**
Dependent Variable: ***School Grade attainment***	**Coefficient**	**Marginal effect**	**Marginal Effect**	**Marginal Effect**
Eastern	0.357	-0.112	0.039**	0.074
	(0.308)	(0.088)	(0.016)	(0.073)
Nyanza	0.358	-0.114	0.041	0.072
	(0.426)	(0.124)	(0.027)	(0.098)
Rift Valley	0.513	-0.158*	0.048***	0.109
	(0.353)	(0.096)	(0.013)	(0.089)
Western	0.430	-0.134	0.043***	0.090
	(0.381)	(0.106)	(0.016)	(0.094)
North Eastern	0.340	-0.105	0.034**	0.072
	(0.479)	(0.132)	(0.016)	(0.118)
Violence	0.351***	-0.111***	0.040***	0.071**
	(0.123)	(0.036)	(0.011)	(0.029)
Any HIV patient	0.118	-0.039	0.018	0.022
	(0.115)	(0.037)	(0.015)	(0.022)
Polygamy	-0.003	0.001	-0.000	-0.001
	(0.117)	(0.040)	(0.020)	(0.020)
Female circumcision	0.133	-0.044	0.020	0.025
	(0.101)	(0.033)	(0.014)	(0.020)
λ	2.079*	-0.713*	0.351*	0.363*
	(1.178)	(0.406)	(0.206)	(0.206)
τ_{1E}	-9.862			
	(9.308)			
τ_{2E}	-8.026			
	(9.311)			

*significant at 10%; **significant at 5%; ***significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)
Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 3.20. **IV Regression Estimates for Rate of Progression during Primary and Secondary School using Age-Qualification.**

Dependent Variable: ***Rate of Progression***	**A: Overall** Number of observations = 3,546 F(35, 1,629) = 28.40	**B: Rural** Number of observations = 2,802 F(33, 1,257) = 20.04	**C: Urban** Number of observations = 744 F(34, 371) = 8.03	**D: Female** Number of observations = 1,746 F(27, 1,137) = 15.73	**E: Male** Number of observations = 1,800 F(28, 1,163) = 23.95
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
Age	-0.118***	-0.112***	-0.132***	-0.109***	-0.125***
	(0.008)	(0.009)	(0.016)	(0.011)	(0.011)
Age squared	0.005***	0.004***	0.005***	0.004***	0.005***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Male	-0.030*	-0.046**	0.022		
	(0.017)	(0.020)	(0.031)		
Male pre-primary aged siblings	-0.033***	-0.035***	-0.020	-0.037***	-0.052***
	(0.008)	(0.009)	(0.018)	(0.009)	(0.010)
Male x Male pre-primary aged siblings	-0.018*	-0.021*	-0.008		
	(0.010)	(0.011)	(0.021)		
Male primary school aged siblings	-0.009	-0.006	-0.023	-0.010	0.004
	(0.006)	(0.007)	(0.015)	(0.006)	(0.006)
Male x Male primary school aged siblings	0.010	0.010	0.019		
	(0.008)	(0.009)	(0.018)		
Male secondary school aged siblings	-0.009	-0.016	-0.017	-0.010	-0.005
	(0.010)	(0.011)	(0.020)	(0.010)	(0.010)
Male x Male secondary school aged siblings	0.004	0.014	-0.004		
	(0.011)	(0.013)	(0.021)		
Female pre-primary aged siblings	-0.018**	-0.021**	0.004	-0.021**	-0.007
	(0.008)	(0.010)	(0.019)	(0.009)	(0.008)
Male x Female pre-primary aged siblings	0.010	0.014	-0.013		
	(0.009)	(0.011)	(0.022)		
Female primary school aged siblings	0.005	0.003	0.015	0.004	-0.006
	(0.006)	(0.008)	(0.013)	(0.007)	(0.007)
Male x Female primary school aged siblings	-0.011	-0.012	-0.013		
	(0.007)	(0.008)	(0.013)		
Female secondary school aged siblings	-0.012	-0.020	0.018	-0.013	-0.017
	(0.010)	(0.012)	(0.016)	(0.012)	(0.012)
Male x Female secondary school aged siblings	-0.006	-0.003	-0.017		
	(0.012)	(0.013)	(0.020)		
Deceased mother	-0.063**	-0.084**	-0.022	-0.059	-0.077*
	(0.029)	(0.039)	(0.042)	(0.039)	(0.044)
Deceased father	-0.003	-0.003	0.050	-0.007	-0.000
	(0.017)	(0.019)	(0.035)	(0.023)	(0.020)

Table 3.20: ***Continued***

Dependent Variable: ***Rate of Progression***	**A: Overall**	**B: Rural**	**C: Urban**	**D: Female**	**E: Male**
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
Both parents deceased	0.015	0.003	-0.025	0.008	0.028
	(0.043)	(0.051)	(0.075)	(0.056)	(0.061)
Protestant & other Christianity	0.005	0.000	0.026	0.003	0.002
	(0.011)	(0.013)	(0.024)	(0.015)	(0.014)
Muslim	-0.045*	-0.007	-0.124***	-0.050	-0.035
	(0.025)	(0.030)	(0.039)	(0.034)	(0.031)
Wealth continuous	0.049***	0.080***	0.019**	0.052***	0.048***
	(0.008)	(0.013)	(0.008)	(0.010)	(0.010)
Head education	0.180***	0.170***	0.208***	0.147***	0.198***
	(0.020)	(0.023)	(0.041)	(0.034)	(0.024)
Urban	0.056***			0.034	0.078***
	(0.018)			(0.023)	(0.021)
Central	0.101***		0.011	0.105***	0.080**
	(0.028)		(0.037)	(0.037)	(0.034)
Coast	-0.004	-0.112***	0.033	0.027	-0.051
	(0.036)	(0.028)	(0.048)	(0.050)	(0.044)
Eastern	0.041	-0.045**	0.029	0.052	0.013
	(0.031)	(0.018)	(0.044)	(0.041)	(0.038)
Nyanza	0.018	-0.072**	0.022	0.037	-0.028
	(0.041)	(0.031)	(0.066)	(0.056)	(0.053)
Rift Valley	0.026	-0.071***	-0.001	0.029	-0.002
	(0.035)	(0.022)	(0.052)	(0.047)	(0.043)
Western	0.013	-0.071***	-0.039	0.029	-0.028
	(0.038)	(0.025)	(0.057)	(0.051)	(0.047)
North Eastern	0.070	-0.033	0.036	0.094	0.027
	(0.050)	(0.046)	(0.075)	(0.076)	(0.057)
Violence	-0.030**	-0.035**	-0.010	-0.044**	-0.023
	(0.013)	(0.016)	(0.030)	(0.018)	(0.017)
Any HIV patient	0.005	0.012	-0.042*	-0.002	0.008
	(0.014)	(0.018)	(0.023)	(0.017)	(0.017)
Polygamy	0.019	0.015	0.057*	0.028	0.015
	(0.015)	(0.017)	(0.033)	(0.022)	(0.018)
Female circumcision	-0.003	0.006	-0.019		-0.006
	(0.012)	(0.014)	(0.026)		(0.013)
λ	-0.190	-0.258	0.103	-0.237	-0.227
	(0.125)	(0.179)	(0.193)	(0.180)	(0.159)
Constant	1.275***	1.390***	1.291***	1.274***	1.293***
	(0.074)	(0.077)	(0.140)	(0.103)	(0.100)

* significant at 10%; ** significant at 5%; *** significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)

Model Set VI: Allowing for Endogenous Wealth (Continuous) and Imputed values for HIV individuals in the missing HIV variable.[43]

Table 3.21. IV Estimates for School Enrolment using Age-Qualification

	3.21a) Dependent Variable: *Primary School Attendance*		3.21b) Dependent Variable: *Secondary School Attendance*	
	Number of observations = 8,209		Number of observations = 2,867	
	Wald χ^2 = 1,076.43		Wald χ^2 = 596.69	
	Prob > χ^2 = 0.000		Prob > χ^2 = 0.000	
	Degrees of freedom = 36		Degrees of freedom = 33	
	Coefficient	**Marginal effect**	**Coefficient**	**Marginal effect**
Age	0.133*	0.013	0.647	0.250
	(0.081)	(0.008)	(0.947)	(0.366)
Age squared	-0.015***	-0.002***	-0.032	-0.012
	(0.004)	(0.000)	(0.029)	(0.011)
Male	0.187	0.019	0.969***	0.357***
	(0.123)	(0.012)	(0.195)	(0.066)
Male pre-primary aged siblings	0.052	0.005	-0.201***	-0.077***
	(0.047)	(0.005)	(0.060)	(0.023)
Male x Male pre-primary aged siblings	-0.082	-0.008	0.193***	0.075**
	(0.053)	(0.005)	(0.086)	(0.033)
Male primary school aged siblings	-0.054*	-0.005*	0.165***	0.064***
	(0.032)	(0.003)	(0.045)	(0.017)
Male x Male primary school aged siblings	0.058	0.006	-0.139**	-0.054**
	(0.048)	(0.005)	(0.062)	(0.024)
Male secondary school aged siblings	0.132**	0.013*	0.121	0.047
	(0.057)	(0.006)	(0.075)	(0.029)
Male x Male secondary school aged siblings	-0.054	-0.005	-0.146	-0.057
	(0.081)	(0.008)	(0.120)	(0.047)
Female pre-primary aged siblings	0.057	0.006	-0.143**	-0.055**
	(0.047)	(0.005)	(0.060)	(0.023)
Male x Female pre-primary aged siblings	-0.012	-0.001	0.255***	0.099***
	(0.059)	(0.006)	(0.091)	(0.035)
Female primary school aged siblings	-0.006	-0.001	0.162***	0.062***
	(0.035)	(0.003)	(0.050)	(0.019)
Male x Female primary school aged siblings	-0.022	-0.002	-0.039	-0.015
	(0.046)	(0.005)	(0.070)	(0.027)
Female secondary school aged siblings	0.077	0.008	0.287***	0.111***
	(0.085)	(0.009)	(0.089)	(0.034)
Male x Female secondary school aged siblings	-0.001	-0.000	-0.532***	-0.205***
	(0.098)	(0.010)	(0.126)	(0.048)
Deceased mother	-0.034	-0.003		
	(0.173)	(0.018)		
Deceased father	0.196*	0.017*		
	(0.104)	(0.010)		
Both parents deceased	-0.687***	-0.113*		
	(0.242)	(0.059)		

[43] ICE –imputation by chain equations (for univariate) is used to impute the missing observations for any household individual who is HIV positive. PCA –Principal Component Analysis is used to obtain a continuous wealth variable given the household characteristics describing the wealth status of the household.

Table 3.21: ***Continued***

	3.21a) Dependent Variable: *Primary School Attendance*		3.21b) Dependent Variable: *Secondary School Attendance*	
	Coefficient	**Marginal effect**	**Coefficient**	**Marginal effect**
Protestant & other Christianity	0.329***	0.034***	-0.013	-0.005
	(0.069)	(0.008)	(0.068)	(0.026)
Muslim	-0.178	-0.019	-0.136	-0.053
	(0.123)	(0.016)	(0.163)	(0.064)
Wealth (Continuous)	0.464***	0.046***	-0.036	-0.014
	(0.079)	(0.016)	(0.106)	(0.041)
Head education	2.022***	0.402***	2.884***	0.701***
	(0.190)	(0.022)	(0.230)	(0.013)
Urban	-0.427***	-0.052*	-0.291	-0.114
	(0.143)	(0.028)	(0.200)	(0.079)
Central	1.383***	0.069***	-0.051	-0.020
	(0.241)	(0.018)	(0.223)	(0.08668
Coast	1.450***	0.065***	-0.104	-0.040
	(0.266)	(0.018)	(0.271)	(0.107)
Eastern	1.597***	0.071***	0.056	0.021
	(0.271)	(0.019)	(0.274)	(0.104)
Nyanza	1.828***	0.075***	0.239	0.090
	(0.255)	(0.019)	(0.283)	(0.103)
Rift Valley	1.244***	0.073***	0.032	0.012
	(0.264)	(0.022)	(0.266)	(0.102)
Western	1.639***	0.070***	0.380	0.140
	(0.259)	(0.018)	(0.289)	(0.100)
North Eastern	1.180***	0.057***	0.224	0.084
	(0.380)	(0.019)	(0.427)	(0.154)
Violence	0.052	0.005	-0.173*	-0.068*
	(0.065)	(0.006)	(0.098)	(0.038)
Any HIV patient (imputed)	0.051	0.005	-0.205**	-0.081**
	(0.100)	(0.009)	(0.083)	(0.033)
Polygamy	-0.171**	-0.019**	-0.065	-0.025
	(0.077)	(0.009)	(0.122)	(0.048)
Female circumcision	-0.081	-0.008	-0.060	-0.023
	(0.067)	(0.006)	(0.080)	(0.031)
Constant	-1.035**		-4.758	
	(0.495)		(7.799)	
$\ln\sigma$	0.204***		0.285***	
	(0.021)		(0.025)	
atanh ρ	-0.542***		-0.000	
	(0.136)		(0.157)	
σ	1.227		1.329	
	(0.026)		(0.033)	
ρ	-0.495		-0.000	
	(0.103)		(0.157)	
	Wald test of exogeneity (atanh ρ = 0):		Wald test of exogeneity (atanh ρ = 0):	
	χ^2_1 = 15.89 Prob > χ^2 = 0.0001		χ^2_1 = 0.00 Prob > χ^2 = 0.9986	

*significant at 10%; **significant at 5%; ***significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)
Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 3.22. Ordered Probit Estimates for School Attainment using Age-Qualification

Number of obs = 2,867 Prob > χ^2 = 0.000		**Marginal effects after ordered probit**		
Wald χ^2 = 347.87 Degrees of freedom = 34		**No schooling**	**Some Primary**	**Complete Primary**
Dependent Variable: *School Grade attainment*	**Coefficient**	**Marginal effect**	**Marginal Effect**	**Marginal Effect**
Age	-0.708	0.254	-0.141	-0.113
	(0.738)	(0.265)	(0.147)	(0.119)
Age squared	0.020	-0.007	0.004	0.003
	(0.023)	(0.008)	(0.004)	(0.004)
Male	-0.426***	0.153***	-0.086***	-0.067***
	(0.136)	(0.049)	(0.028)	(0.021)
Male pre-primary aged siblings	0.080	-0.029	0.016	0.013
	(0.054)	(0.019)	(0.011)	(0.009)
Male x Male pre-primary aged siblings	-0.049	0.017	-0.010	-0.008
	(0.067)	(0.024)	(0.013)	(0.011)
Male primary school aged siblings	-0.075**	0.027**	-0.015**	-0.012**
	(0.033)	(0.012)	(0.007)	(0.005)
Male x Male primary school aged siblings	0.109**	-0.039**	0.022**	0.018**
	(0.043)	(0.015)	(0.009)	(0.007)
Male secondary school aged siblings	-0.073	0.026	-0.015	-0.012
	(0.057)	(0.020)	(0.011)	(0.009)
Male x Male secondary school aged siblings	-0.021	0.008	-0.004	-0.003
	(0.079)	(0.028)	(0.016)	(0.013)
Female pre-primary aged siblings	0.081*	-0.029*	0.016*	0.013*
	(0.046)	(0.016)	(0.009)	(0.007)
Male x Female pre-primary aged siblings	-0.076	0.027	-0.015	-0.012
	(0.062)	(0.022)	(0.012)	(0.010)
Female primary school aged siblings	-0.050	0.018	-0.010	-0.008
	(0.038)	(0.014)	(0.008)	(0.006)
Male x Female primary school aged siblings	0.067	-0.024	0.013	0.011
	(0.050)	(0.018)	(0.010)	(0.008)
Female secondary school aged siblings	-0.172***	0.062***	-0.034***	-0.028***
	(0.061)	(0.022)	(0.012)	(0.010)
Male x Female secondary school aged siblings	0.272***	-0.098***	0.054***	0.044***
	(0.089)	(0.032)	(0.018)	(0.014)
Protestant & other Christianity	-0.020	0.007	-0.004	-0.003
	(0.053)	(0.019)	(0.010)	(0.009)
Muslim	0.056	-0.020	0.011	0.009
	(0.130)	(0.046)	(0.024)	(0.0219)
Wealth	-0.119**	0.043**	-0.024**	-0.019**
	(0.052)	(0.019)	(0.011)	(0.008)
Wealth Residual	0.017	-0.006	0.003	0.003
	(0.054)	(0.019)	(0.011)	(0.009)
Head education	3.413***	-0.788***	0.628***	0.159***
	(0.281)	(0.009)	(0.011)	(0.008)
Urban	0.163	-0.057	0.030*	0.028
	(0.109)	(0.037)	(0.018)	(0.020)
Central	-0.351**	0.132**	-0.084*	-0.048**
	(0.168)	(0.065)	(0.046)	(0.019)
Coast	-0.160	0.059	-0.035	-0.024
	(0.189)	(0.071)	(0.046)	(0.025)
Eastern	-0.143	0.052	-0.031	-0.021
	(0.179)	(0.067)	(0.042)	(0.025)

Table 3.22: ***Continued***

		Marginal effects after ordered probit		
		No schooling	**Some Primary**	**Complete Primary**
Dependent Variable: ***School Grade attainment***	**Coefficient**	**Marginal effect**	**Marginal Effect**	**Marginal Effect**
Nyanza	-0.571***	0.217***	-0.148***	-0.070***
	(0.182)	(0.071)	(0.055)	(0.017)
Rift Valley	-0.167	0.061	-0.036	-0.028
	(0.178)	(0.067)	(0.042)	(0.024)
Western	-0.368**	0.139**	-0.089*	-0.049**
	(0.181)	(0.070)	(0.051)	(0.020)
North Eastern	-0.666**	0.257**	-0.186**	-0.072***
	(0.262)	(0.102)	(0.085)	(0.018)
Violence	0.095	-0.033	0.018	0.016
	(0.069)	(0.024)	(0.012)	(0.012)
Any HIV patient	0.075	-0.026	0.014	0.012
	(0.074)	(0.026)	(0.013)	(0.013)
Polygamy	-0.027	0.010	-0.005	-0.004
	(0.079)	(0.029)	(0.016)	(0.012)
Female circumcision	0.049	-0.017	0.010	0.008
	(0.057)	(0.020)	(0.011)	(0.009)
τ_{1E}	-4.120			
	(6.044)			
τ_{2E}	-2.309			
	(6.045)			

*significant at 10%; **significant at 5%; ***significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)
Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 3.23. IV Estimates for Rate of Progression during Primary and Secondary School using Age-Qualification

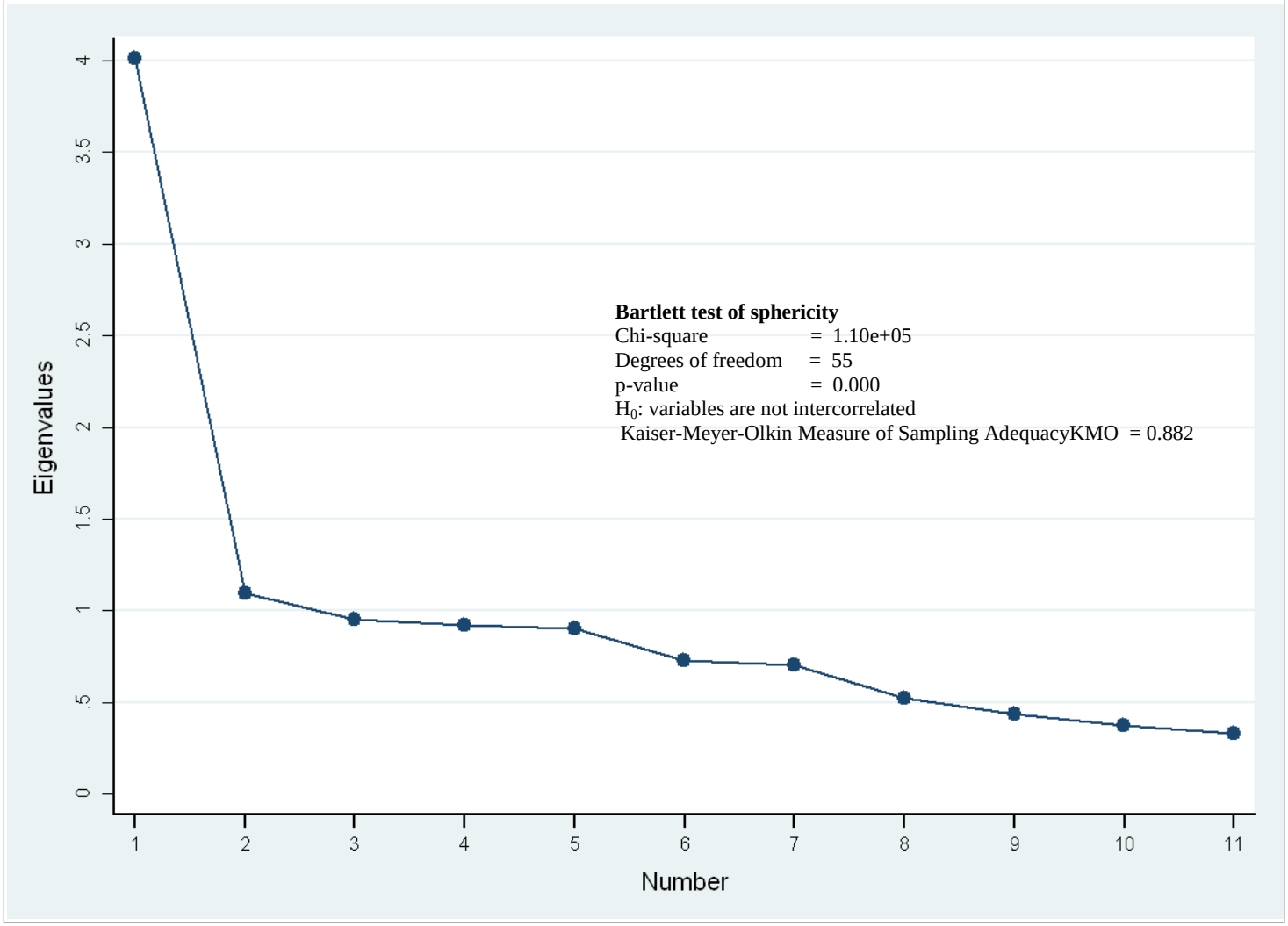
Dependent Variable: ***Rate of Progression***	**A: Overall** Number of observations = 8,030 F(36, 3,701) = 58.95	**B: Rural** Number of observations = 6,225 F(34, 2,767) = 44.17	**C: Urban** Number of observations = 1,805 F(35, 933) = 15.32	**D: Female** Number of observations = 3,911 F(29, 2,557) = 36.03	**E: Male** Number of observations = 4,119 F(29, 2,649) = 41.09
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
Age	-0.122***	-0.124***	-0.111***	-0.116***	-0.126***
	(0.005)	(0.006)	(0.011)	(0.007)	(0.008)
Age squared	0.005***	0.005***	0.004***	0.005***	0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	-0.041***	-0.050***	-0.034		
	(0.012)	(0.013)	(0.022)		
Male pre-primary aged siblings	-0.023***	-0.024***	-0.030**	-0.027***	-0.026***
	(0.005)	(0.006)	(0.012)	(0.005)	(0.006)
Male x Male pre-primary aged siblings	-0.005	-0.004	-0.002		
	(0.006)	(0.007)	(0.014)		
Male primary school aged siblings	-0.008*	-0.005	-0.023**	-0.008*	0.002
	(0.004)	(0.005)	(0.010)	(0.004)	(0.005)
Male x male primary school aged siblings	0.009*	0.006	0.036***		
	(0.005)	(0.006)	(0.011)		
Male secondary school aged siblings	-0.006	-0.004	-0.033**	-0.008	-0.009
	(0.006)	(0.007)	(0.014)	(0.006)	(0.007)
Male x Male secondary school aged siblings	-0.005	-0.005	0.026		
	(0.008)	(0.008)	(0.017)		
Female pre-primary aged siblings	-0.008	-0.011*	0.008	-0.010*	-0.001
	(0.005)	(0.006)	(0.010)	(0.005)	(0.006)
Male x Female pre-primary aged siblings	0.005	0.009	-0.006		
	(0.006)	(0.007)	(0.013)		
Female primary school aged siblings	0.003	0.001	0.004	0.003	0.001
	(0.004)	(0.004)	(0.009)	(0.004)	(0.004)
Male x Female primary school aged siblings	-0.002	-0.001	-0.005		
	(0.005)	(0.005)	(0.011)		
Female secondary school aged siblings	-0.006	-0.012	0.004	-0.007	0.001
	(0.006)	(0.008)	(0.009)	(0.006)	(0.009)
Male x Female secondary school aged siblings	0.005	0.006	-0.006		
	(0.008)	(0.009)	(0.015)		
Deceased mother	-0.027	-0.014	-0.057	-0.021	-0.030
	(0.021)	(0.024)	(0.035)	(0.027)	(0.032)
Deceased father	-0.006	-0.016	0.034	-0.009	0.001
	(0.012)	(0.014)	(0.023)	(0.016)	(0.016)

Table 3.23: ***Continued***

Dependent Variable: *Rate of Progression*	A: Overall	B: Rural	C: Urban	D: Female	E: Male
	Coefficient	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
Both parents deceased	0.013	0.004	0.002	0.009	0.011
	(0.031)	(0.035)	(0.053)	(0.041)	(0.045)
Protestant & other Christianity	0.009	0.010	0.009	0.014*	0.004
	(0.007)	(0.008)	(0.013)	(0.008)	(0.009)
Muslim	-0.056***	-0.012	-0.113***	-0.051**	-0.061***
	(0.016)	(0.020)	(0.023)	(0.021)	(0.020)
Wealth (Continuous)	0.063***	0.085***	0.029***	0.055***	0.072***
	(0.006)	(0.008)	(0.007)	(0.007)	(0.009)
Head education	0.187***	0.174***	0.214***	0.168***	0.202***
	(0.013)	(0.014)	(0.025)	(0.020)	(0.017)
Urban	0.005			0.002	0.005
	(0.013)			(0.016)	(0.018)
Central	0.173***	0.028	0.030	0.165***	0.182***
	(0.020)	(0.029)	(0.026)	(0.023)	(0.028)
Coast	0.076***	-0.071***	0.028	0.078***	0.078**
	(0.023)	(0.025)	(0.022)	(0.029)	(0.032)
Eastern	0.109***	-0.018	0.051*	0.102***	0.115***
	(0.022)	(0.025)	(0.030)	(0.027)	(0.031)
Nyanza	0.117***	-0.011	0.026	0.104***	0.130***
	(0.022)	(0.025)	(0.024)	(0.026)	(0.031)
Rift Valley	0.103***	-0.029	0.002	0.080***	0.125***
	(0.021)	(0.025)	(0.027)	(0.025)	(0.029)
Western	0.093***	-0.037	0.002	0.087***	0.100***
	(0.022)	(0.025)	(0.027)	(0.027)	(0.031)
North Eastern	0.170***		0.062	0.120***	0.209***
	(0.034)		(0.042)	(0.045)	(0.044)
Violence	-0.015*	-0.020**	-0.025	-0.027**	-0.003
	(0.008)	(0.009)	(0.018)	(0.011)	(0.010)
Any HIV patient	0.004	0.005	-0.009	0.001	0.007
	(0.008)	(0.007)	(0.013)	(0.011)	(0.011)
Polygamy	-0.004	-0.012	-0.009	-0.002	-0.006
	(0.011)	(0.011)	(0.025)	(0.015)	(0.013)
Female circumcision	0.003	0.006	-0.002	0.008	-0.003
	(0.007)	(0.008)	(0.016)	(0.009)	(0.010)
Constant	1.178***	1.334***	1.213***	1.174***	1.138***
	(0.039)	(0.043)	(0.066)	(0.049)	(0.057)

* significant at 10%; ** significant at 5%; *** significant at 1%
Robust standard errors are in parentheses (adjusted for clustering on household level)

Figure 3.1. **Scree Plot of eigenvalues after PCA for household assets**



Appendix 3.A

Equation (3.12) can be transformed as

$$\left(SchAtt_i^* - \gamma_i W_i^*\right) = X_i'\beta_1 + u_{1i}$$
$$W_i^* = X_{Wi}'\beta_2 + u_{2i} \qquad (A3.1)$$

In matrix form, equation (A3.1) can be written as follows:

$$\Gamma Y^* = \mathrm{B}X + \mathrm{E}, \quad \text{with } \mathrm{E} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \Sigma\right) \text{ and } \Sigma = \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix} \qquad (A3.2)$$

where:

$$\Gamma = \begin{pmatrix}1 & 0\\ -\gamma_i & 1\end{pmatrix};\ Y^* = \begin{pmatrix}SchAtt_i^*\\ W_i^*\end{pmatrix};\ X = \begin{pmatrix}X_i\\ X_{Wi}\end{pmatrix};\ \mathrm{B} = \begin{pmatrix}\beta_1 & 0\\ 0 & \beta_2\end{pmatrix} \text{ and } \mathrm{E} = \begin{pmatrix}u_{1i}\\ u_{2i}\end{pmatrix} \qquad (A3.3)$$

By multiplying equation (A3.2) by Γ^{-1} gives

$$Y^* = \Lambda X + \Psi \qquad (A3.4)$$

where $\Lambda = \Gamma^{-1}\mathrm{B}$, $\Psi = \Gamma^{-1}\mathrm{E}$ and $\Psi \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \Xi\right)$

In this case,

$$\Xi = \left(\Gamma^{-1}\right)\Sigma\left(\Gamma^{-1}\right)^T$$
$$= \begin{pmatrix}1 & 0\\ \gamma_i & 1\end{pmatrix}\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\begin{pmatrix}1 & \gamma_i\\ 0 & 1\end{pmatrix}$$
$$= \begin{pmatrix}1 & \rho+\gamma_i\\ \rho+\gamma_i & \gamma_i^2 + 2\gamma_i\rho + 1\end{pmatrix} \qquad (A3.5)$$

Define a matrix Ω with the principal diagonal of inverse squared root[44] of the terms in the principal diagonal of Ξ and zero elsewhere given by: $\Omega = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sqrt{\gamma_i^2 + 2\gamma_i\rho + 1}} \end{pmatrix}$.

Multiplying equation (A3.5) by this matrix Ω we obtain the equation:

$$\Upsilon Y^* = \Upsilon \Xi X + \Upsilon \mathrm{E} \qquad (A3.6)$$

where the error term transforms to:

$$\Upsilon \mathrm{E} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Upsilon \Xi \Upsilon\right), \text{ with}$$

$$\Upsilon \Xi \Upsilon = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sqrt{\gamma_i^2 + 2\gamma_i\rho + 1}} \end{pmatrix} \begin{pmatrix} 1 & \rho + \gamma_i \\ \rho + \gamma_i & \gamma_i^2 + 2\gamma_i\rho + 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sqrt{\gamma_i^2 + 2\gamma_i\rho + 1}} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \frac{\rho + \gamma_i}{\sqrt{\gamma_i^2 + 2\gamma_i\rho + 1}} \\ \frac{\rho + \gamma_i}{\sqrt{\gamma_i^2 + 2\gamma_i\rho + 1}} & 1 \end{pmatrix}$$

$$\equiv \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \qquad (A3.7)$$

where $\rho = \frac{\rho + \gamma_i}{\sqrt{\gamma_i^2 + 2\gamma_i\rho + 1}}$

[44] See Sherif (1991)

Appendix 3.B: Variable Glossary

Dependent Variables		
Primary School Attendance	Dummy = 1 if child currently attends school and aged 6 - 14 omitting above secondary	
Secondary School Attendance	Dummy = 1 if child currently attends school and aged 15 - 18 omitting above secondary	
School Grade attainment	Grade attained: 0 = No Schooling; 1 = Some but not all primary schooling; and 2 = Completed Primary School	
Grade	The number of grades a child has passed	
Bivariate second equation		
Household Wealth for Bivariate second equation	Household Wealth index (Quintile)	
Household member agree for HIV test *(for selection equation)*	Dummy = 1 if the individual had agreed to be tested for HIV given was selected	
Independent Variables	**Description**	**Base** (for dummy variables) **Units** (for count or continuous variable)
Child characteristics		
Male	Gender dummy = 1 if child is male and aged 0 - 18 years	Female
Male Pre-Primary Aged Siblings Male Primary School Aged Siblings Male Secondary School Aged Siblings	Pre-Primary School Age is less than 6 years Primary School Aged is between 6 - 14 years Secondary School Aged is between 15 – 18 years	Number of siblings[45]
Female Pre-Primary Aged Siblings Female Primary School Aged Siblings Female Secondary School Aged Siblings	Pre-Primary School Age is less than 6 years Primary School Aged is between 6 - 14 years Secondary School Aged is between 15 – 18 years	
Interactions	Interaction of male child and the above siblings according to age group and gender	
Deceased mother	Dummy = 1 if the mother of the child in the household is deceased	Mother alive
Deceased father	Dummy = 1 if the father of the child in	Father Alive

[45] Count variable giving the number of male siblings of the child. Children who have completed secondary school are not included regardless their age

	the household is deceased	
Both Parents Deceased	Dummy = 1 if both parents of the child in the household are deceased	Both Parents Alive
Female circumcision	Dummy = 1 if the female child is circumcised	No circumcised
Household characteristics		
Religion		
Number of children aged 0 and 5 years	Number of children in the household aged 0 – 5 years old	Number of children
Number of children aged 6 and 14 years	Number of children in the household aged 6 – 14 years old	
Number of children aged 15 and 18 years	Number of children in the household aged 15 – 18 years old	
Catholic	Dummy = 1 if the individual in the household belongs to Catholic religion	Not Catholic
Protestant and other Christianity	Dummy = 1 if the individual household belongs to Protestant/Other Christian religion	Not Protestant and Not other Christian
Muslim	Dummy = 1 if the individual household belongs to Islamic religion	Not Muslim
No Religion	Dummy = 1 if the household belong to no religion	Religious
Other religion	Dummy = 1 if the individual household belong to OTHER religion	Not other religion
Wealth Index		
Poorest household Poorer household Middle household Richer household Richest household	Dummy = 1 if the individual is in the classified household wealth named	Poorest Household
Wealth, Continuous	Predicted score from the first principle component from PCA for household assets	
Wealth Continuous residuals	Residuals after regressing Wealth (Continuous) variable on selected explanatory variables	
Residence/Region		
Urban	Dummy = 1 if the individual is in the household residing in urban area	Rural

Nairobi Central Coast Eastern Nyanza Rift Valley Western North Eastern	Dummy = 1 if the individual is in the household residing in named Province	Nairobi
Child's/Individual Characteristics		
Age Age squared	Child's age. Age squared is included to allow for nonlinear effect of age.	Years
Household head age Household age squared	Household head age and it's squared variable to allow for nonlinearity effect	
Head education	Dummy = 1 if the household head has some primary school education	No Primary education
Mother hurt child	Dummy = 1 if the child has ever been physically hurt by the Mother	Not hurt by mother
Father hurt child	Dummy = 1 if the child has ever been physically hurt by the Father	Not hurt by father
Teacher hurt child	Dummy = 1 if the child has ever been physically hurt by the Teacher	Not hurt by teacher
Any HIV patient	Dummy = 1 if any household member is HIV positive	Not HIV patient
Husband live away	Dummy = 1 if the husband of the household stays elsewhere	Live with Family
Violence	Dummy = 1 if the wife of the household experienced any violence	No Violence
Polygamy	Dummy = 1 if woman is married with other wife (ves) and 0 if only wife or single	Monogamy
Female circumcision	Dummy = 1 if the female child is circumcised	No circumcised
Union	Dummy = 1 if the individual had more than one unions	One Union
Working away	Dummy = 1 if individual work away from home	Working at Home
Household member consent	Dummy = 1 if defined by household, given that any person in the selected household agreed for HIV/AIDS testing.	
Tribe		
Embu	Dummy = 1 if the individual is from Embu tribe and 0 if other tribe	
Kalenjin	Dummy = 1 if the individual is from Kalenjin tribe and 0 if other tribe	

Kamba	Dummy = 1 if the individual is from Kamba tribe and 0 if other tribe	
Kikuyu	Dummy = 1 if the individual is from Kikuyu tribe and 0 if other tribe	
Kisii	Dummy = 1 if the individual is from Kisii tribe and 0 if other tribe	
Luhya	Dummy = 1 if the individual is from Luhya tribe and 0 if other tribe	
Luo	Dummy = 1 if the individual is from Luo tribe and 0 if other tribe	
Masai	Dummy = 1 if the individual is from Masai tribe and 0 if other tribe	
Meru	Dummy = 1 if the individual is from Meru tribe and 0 if other tribe	
Mijikenda/Swahili	Dummy = 1 if the individual is from Mijikenda/Swahili tribe and 0 if other tribe	
Somali	Dummy = 1 if the individual is from Somali tribe and 0 if other tribe	
Taita	Dummy = 1 if the individual is from Taita tribe and 0 if other tribe	
Turkana	Dummy = 1 if the individual is from Turkana tribe and 0 if other tribe	
Kuria	Dummy = 1 if the individual is from Kuria tribe and 0 if other tribe	
Other	Dummy = 1 if the individual is from Other tribes and 0 if other tribe	
Occupation/Work type		
Professional, technical or Managerial	Dummy = 1 if an individual's occupation is Professional, technical or Managerial and 0 otherwise	
Clerical	Dummy = 1 if an individual's occupation is Clerical and 0 otherwise	
Sales	Dummy = 1 if an individual's occupation is Sales and 0 otherwise	
Agricultural - self employed	Dummy = 1 if an individual's occupation is Agricultural self-employed and 0 otherwise	
Household & Domestic	Dummy = 1 if an individual's occupation is Household and Domestic work and 0 otherwise	
Services	Dummy = 1 if an individual's occupation is Services and 0 otherwise	
Skilled Manual	Dummy = 1 if an individual's occupation is Skilled Manual and 0otherwise	
Unskilled Manual	Dummy = 1 if an individual's occupation is Unskilled Manual and 0 otherwise	

Unknown	Dummy = 1 if individual's occupation is unknown or Not working and 0 otherwise	
Seasonally Working	Dummy = 1 if an individual work is seasonal and 0 if Otherwise	
Permanently Working	Dummy = 1 if an individual work is permanent and 0 Otherwise	
Not Working	Dummy = 1 if an individual is Not working or works occasionally and 0 Otherwise	
σ [46]	The standard error of the residuals of the child's schooling equation	
ρ [47]	Correlation between the residuals of wealth (or selection) equation and child schooling equation	
λ [48]	Inverse Mills Ratio: Selectivity bias correction factor computed from the estimated household member testing HIV/AIDS positive	
τ_{1E}	Estimated threshold parameter for the cut point between No Education and Some Primary Education	
τ_{2E}	Estimated threshold parameter for the cut point between Some Primary Education and Complete Primary Education	
τ_{1W}	Estimated threshold parameter for the cut point between poorest and poorer household	
τ_{2W}	Estimated threshold parameter for the cut point between poorer and middle household	
τ_{3W}	Estimated threshold parameter for the cut point between middle and richer household	
τ_{4W}	Estimated threshold parameter for the cut point between richer and richest household	

[46] Actually the log of σ, i.e $\ln\sigma$ because σ is not directly estimated

[47] Actually the inverse hyperbolic tangent of ρ, i.e. $\text{atanh}\rho = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$. because ρ is not directly estimated

[48] Which is $\rho * \sigma$

CHAPTER 4

IMPACT OF HOUSEHOLD CHARACTERISTICS ON CHILD HEALTH

4.1 Introduction

This chapter deals with children's health. When we talk about child health, we need to be aware of the factors which enhance or hinder good health. When a child is born, his or her survival can be impacted by how and where they live and the characteristics of the people or community around them. According to the WHO definition, a healthy child could be considered as one that grows with complete physical, mental and social well-being without any effect of diseases or infirmity of any kind (WHO, 1946). This definition has not been amended since this time.

As the child grows, there are always a variety of risks to their survival. Infant and child mortality, specifically during the first year of life and during the following 4 years of life, are important indicators for mortality conditions, health progress and indeed the overall social and economic well-being of a country (Macassa, *et al.*, 2003). Rahman (2009), in his study of factors affecting child survival in Bangladesh, points out that mortality rates in the first year and in the subsequent 4 years of life have attracted particular attention because: i) mortality is relatively high in these periods, ii) they have a considerable impact on the average life expectancy and the rate of population growth, iii) they are sensitive to environmental and sanitary conditions and iv) the levels of

infant and child mortality are useful indicators of the state of health of a society or a country.

Mortality rates among children under the age of five remain strikingly high throughout the majority of sub-Saharan Africa compared to most other areas of the world. Despite action plans by policy makers, including the Millennium Development Goals (MDGs), there are many reasons why these rates remain. Kenya is part of this region and it is important to determine the major contributing factors towards early child health, specifically infant and child mortality, hence the need for the research carried out in this chapter.

According to WHO statistics on Kenya, the proportion of children who die before age 5 is 12.1%. Just as in every country, the child mortality rate worsened in Kenya shortly after the Second World War (WHO, 2009). This phenomenon has been observed and studied by various researchers. Hill (1992) highlighted the significant deterioration in the infant mortality rate since the 1940s. In particular, the infant mortality rate declined from 184 deaths per 1,000 live births in 1948, to 104 deaths in 1979, to 62 in 1989. However, the situation changed after that where the rates increased to 74 and 77 in 1998 and 2003 respectively. This reversal of trends in mortality decline since mid–1990s was largely attributed to the impact of HIV/AIDS pandemic, widespread poverty levels and the deterioration in levels of health care in the country (Obonyo, Otieno and Muga, 2005; Brass, 1993). It has been a major problem to get accurate information on the cause of deaths of children under five years of age in Kenya, but like most countries in sub-Saharan Africa, the cause of deaths at this early age, is probably dominated by pneumonia, malaria, measles and diarrheal disease, which are estimated to have been

responsible for some 60% of disease burdening the region as of 1990 (Murray and Lopez 1996).

Previous studies on children's heath under five years of age in Kenya have looked into medical, socioeconomic determinants and demographic factors as the causes of infant and child mortality rates. Such studies include that done by McElroy *et al.* (2001) using longitudinal data from 1992 to 1996 in Western Kenya under Asembo Bay Cohort Project (ABCP) where they determined the causes of mortality among young children in western Kenya. Omariba *et al*. (2007), in their study of the determinants of infant and child mortality in Kenya, looked into Hazard ratios of child mortality using the 1998 Kenya Demographic Health Survey (KDHS). Mutunga (2007) using the Kenya Demographic and Health Survey (KDHS) 2003, examined how infant and child mortality is related to the household's environmental and socio-economic characteristics, such as mother's education, source of drinking water, sanitation facility, type of cooking fuels and access to electricity. He used a hazard rate framework to analyse the determinants of child mortality. A recent study by Mustafa and Odimegwu (2008) has looked into rank-ordering of the important factors, still looking into their effect on infant mortality.

The literature reveals that there are at least three frameworks which may be used in studying the determinants of childhood mortality (Mamo, 1993). There is a medical science approach which argues that child mortality is a biological process which results from the incidence of diseases among children, influenced by biomedical factors, such as environmental contamination, injury, personal illness control and dietary intake. The other framework is the social science approach, which studies childhood mortality as an indirect result of socioeconomic factors. Abate (1993) points out that this approach pays

loose attention to the means by which these variables operate and that the unobserved biomedical factors have been shown to cluster within socioeconomic variables.[49] The third framework is the one which was proposed by Mosley and Chen (1984). This framework tries to integrate these two other approaches by showing that socioeconomic factors affect child mortality through biomedical factors which they call intermediate (intervening) variables.

Unfortunately, despite this framework being more promising for child mortality studies, most researchers have not utilised it due to a lack of information on the intermediate variables in developing economies. In this research we will assess determinants of child health using the Mosley and Chen approach. The assessment of child health is essential if child mortality is to be reduced in Kenya and any other part of the world. Determining the key factors responsible for child health will enable mothers and health policy makers to understand in depth the factors important for producing a healthy child and as a result reduce infant and child mortalities. To do this, we are going to model child health by modelling the z-score for weight, height and *bmi* allowing these to be affected by biomedical, socioeconomic and demographic factors, using a cross sectional data set, the Kenyan Demographic and Heath Survey, 2003.

The rest of this chapter will proceed in the following manner: the next section will give the relevant literature on child health assessment. An economic framework will then be presented, followed by presentation of econometric methodology, including the definitions and derivations of the variables and the empirical specification. Thereafter, we will briefly introduce the organisation of the survey and the data and then the

[49] See also Caldwell, 1979; Farah & Preston, 1982; United Nations, 1985; Hobcraft, McDonald & Rutstein, 1984

discussion of our results. The chapter concludes with discussion and concluding remarks.

4.2 Relevant Literature

Pine (1997) has outlined the history of medicine and public health as consisting of four phases; the first phase, spanning from the middle to the late 1800s, was characterized by urbanization and industrialization which led to consideration of living conditions for workers and sanitation and associated to public health reforms. The second phase was between 1880 and 1930 which was characterized by advances in bacteriology and immunology. The third phase was from 1930 through to 1974 and was a therapeutic period. It is in this phase where hospitals became the essential base and focus for medical services with medical treatment becoming more complex. Pine mentions that with the development of vaccines and antibiotics, along with the success of surgical procedures, people began to rely on medical interventions as the source of health. The biomedical approach became paramount, and people began to believe that health was delivered to them by health professionals.

Pine gives the fourth phase as from 1974 to the present. It is in this phase where researchers have embarked on looking into factors which may be responsible for mortalities. There has been development of conceptual models addressing these health matters, for example Dahlgren and Whitehead (1995), who proposed the model showing that the individual is surrounded by lifestyle factors, social and community influences, living and working conditions, and general socioeconomic, cultural, and environmental conditions. Green and Ottoson (1999) integrated an earlier model developed by Lalonde

(1974) into a framework of population health strategies, processes of change, determinants of health, and ultimate social and health outcomes.

There has been a lot of empirical work on child health in Africa. Caldwell (1979), in his study using Nigerian data, tried to link maternal education levels as a factor contributing to child mortality decline in Nigeria. In addition, he used several other socio-economic characteristics in association with child mortality. Caldwell found that a mother's education affects child mortality. The possible reason he gave to explain this effect was that mother's education works through changing feeding and care practices, leading to better health seeking behaviour and by changing the traditional family structure. Similarly, Hobcraft (1993), using DHS data from different countries (sub-Saharan African), found that maternal education was more likely to contribute to child survival by enabling women to have fewer children and utilising pre-natal care and immunising their children. However, in his earlier studies, Hobcraft *et al.* (1984) had found that the effect in sub–Saharan Africa was weak. Woldemicael (2001), using data from the 1995 Eritrea Demographic and Health Survey also looked at the effect of some environmental and socioeconomic factors determining childhood diarrhea in Eritrea. Madise and Diamond (1995) using data for the 1988 Malawi Traditional and Modern Methods of Child Spacing survey, looked into similar factors contributing to child mortality.

In contrast, in their study using 1991/92 Tanzania Demographic Health Surveys, Mturi and Curtis (1995) found that socioeconomic factors such as maternal education, partner's education, urban/rural residence, economic status (wealth index), ethnicity and gender on infant and child mortality differentials had an insignificant impact. Their study showed demographic factors including short birth intervals (less than 2 years), teenage pregnancies (less than 20 years of age) and a previous child's death were

associated with an increase in infant and child mortality in the country. However, more recent studies, such as Hosseinpoor *et al.* (2005), which studied the socioeconomic inequality in infant mortality in Iran and across its provinces, using Iran 2000 DHS data, has again shown the link between socioeconomic and demographic factors and child mortality.

Other studies have used biomedical factors to explain infant child mortality. Mosley and Chen (1984) are an example of early researchers who studied the intermediate biomedical factors affecting child mortality. Their study is referred to extensively in other literature (for example Mutunga, (2007)).

4.2.1 Measuring Child Health

To assess factors which influence a child's health, we first need to decide on a measure of child health. In 2006 the World Health Organization (WHO) released a set of child growth standards for children aged 0 – 60 months which included Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and a body mass index-for age and sex (WHO, 2006). These standards were to replace the National Center for Health Statistics references established in 1977 (Hamill *et al.,* 1977). In addition, in 2007 WHO gave other indicators which are able to be used in assessment of children's health. These included head circumference for age, arm circumference for age, triceps skinfold for age and subscapular skinfold for age (WHO, 2007).

These standards give the distribution of weight, length, height, and body mass index (*BMI*) by age and sex and distribution of weight by length and sex (Van den *et al*., 2009). However, these indicators require statistical transformations to make them

suitable as universal indicators of child's health throughout the world. The standardisation of these indicators involves the construction of z-scores where measures are standardised to have mean zero and variance one by subtracting mean and dividing by standard deviation for the relevant age and sex of each child. (For more detail, see WHO (1995); Onis M, *et al*. (2003)). The WHO Global Database on Child Growth and Malnutrition uses a z-score cut-off point of less than $-2SD$ to classify low weight-for-age, low height-for-age and low weight-for-height as moderate, and less than $-3SD$ to define severe under nutrition. The cut-off point of greater than $+2SD$ classifies high weight-for-height as overweight in children. This usage of -2 as a cut-off implies that 2.3% of the reference population will be classified as malnourished even if they are truly "healthy" individuals with no growth impairment, hence, making 2.3% to be regarded as the baseline or expected prevalence. These indicators have been used by many researchers in modeling child health (for example Onis *et al.* (2004), Kulanga *et al.* (2010)).

In addition to the above standards, there are other standards which have been used for child health assessment which were established in 2000 by the Centers for Disease Control and Prevention (CDC). They are used as growth charts for the United States (Kuczmarski *et al.*, 2002). These standards mostly are for children 2 – 20 years of age. The 1990 British growth charts (United Kingdom version) have also been used by several researchers. The existence of many different standards has led to confusion over which standard should be used as a universal scale. Several studies have been established to compare these various standards and analyse which is the most reliable. Wright *et al*. (2002) has compared these standards with the original Tanner–Whitehouse and the Gairdner–Pearson growth references for weight and height/length, they

concluded that for children under the age of 2 years and over 2 years, the UK1990 standards are suitable reference charts for weight, length, and BMI for age and sex.

It was only recently that a standard definition of overweight and obesity for children was established. This was after the Childhood Obesity Working Group of the International Obesity Taskforce recommended the use of BMI cut-off points to categorize children as normal weight, overweight, or obese based on age, gender, and BMI (Cole *et al*., 2000). Suzanna *et al.* (2004), observes that these cut-off points were developed using large, nationally representative, cross-sectional datasets from six countries: Brazil, Great Britain, Hong Kong, the Netherlands, Singapore, and the United States and they corresponded to equivalent adult BMI cut-off points endorsed by the World Health Organization of $BMI < 25 kg/m^2$ for normal weight, BMI of $25 - 29.99 kg/m^2$ for overweight, and $BMI > 30 kg/m^2$ for obesity.

In this study we are not considering particular health conditions or health shocks, but we are more interested in a child's overall health, as this is more likely to correlate to important socio-economic factors such as household wealth.

Of the measures we consider, height for age is generally regarded as the best indicator of overall, long-term health. Once we control for genetic factors, a child's height will on average, reflect their levels of nutrition over the early years of their life. A child who is particularly short is described as "stunted", a condition that arises from the cumulative effect of chronic malnutrition. Stunting may also be the result of recurrent or chronic illness. BMI is normally used as an indicator of current, short-term nutritional status. A child whose weight is particularly low relative to their height might be described as "wasted", reflecting acute malnutrition. Severe wasting will in turn be associated with

significantly increased risk of mortality. The third measure, weight-for-age is a composite index which reflects the combined effects of both long term nutrition and chronic health conditions (revealed in height) and acute malnutrition or illness (captured by BMI).

Before continuing on to the econometric methodology, a theoretical framework for the relationship between children's health and human capital development will be introduced.

4.3 Theoretical Framework

In this section, a household health production function model will be explored. More specific models and derivation will subsequently be closely examined.

4.3.1 Human Capital and Child health

Human capital is an essential factor in economic growth. Ray (1998) defines human capital as the labour that is skilled in production, labour that can operate sophisticated machinery, labour that can create new ideas and new methods in economic activity. Becker (1975, 1994) broadens this definition of investment in human capital to include: schooling, computer training courses, expenditure on medical care, and even lectures on the virtues of punctuality and honesty. He includes these elements because they raise earnings, improve health and add to a person's good habits over their lifetime. It is with this understanding that economists regard expenditure on education, training, medical care, and so on as investments in human capital. These are collectively called "human capital" because people cannot be separated from their knowledge, skills, health, or values in the way that they can be separated from their financial and physical assets.

Children's health significantly contributes towards human capital and vice versa. Investment in the human capital of children is likely to improve their current welfare and enhance opportunities over their life span. In various research studies conducted on human capital investments, specifically schooling and health, children's health has generally taken priority. This prioritisation acknowledges the impact of a child's health on their schooling achievements. In developing countries, specifically in poverty stricken areas, it is common for children to be sick and malnourished and therefore investment in medical and nutritional health (not only medical attention but also nutritional) will be more likely to have a substantial positive impact on human capital. Investment in health reduces mortality and susceptibility to infectious disease that may impair cognitive development.

4.3.2 A Children's health production function

In economics, it is established that parents are the key decision makers in their child's development specifically in areas such as education and health. An economic model that can be utilized in assessing children's health is essential in understanding the factors which impact children's health. The framework developed by Grossman (1972) will be used to give a basic structure of the health production model. Grossman's Model has been used by various researchers to study children's health (for example Rosenzweig and Schultz, 1983; Grossman and Joyce, 1990). The household production function has been developed within the basic framework created by Becker (1965, 1967). Here, it is assumed that the household can, in part, derive its utility from children being healthy. This reflects the view that health is like other goods and services. The household provides inputs that result in health, educational, cognitive, and socio-emotional

outcomes for their child. Examples of these inputs are the child's nutritional intake, educational resources, use of health services, entertainment, housing, babysitting and non-curriculum education, even household contributions to the local community.

In his model, Grossman assumes education is an environmental input which influences the efficiency of the health production process, in that more educated people are more efficient producers of health (Grossman, 1972). Other researchers such as Leigh (1990) and Kutty (2000) have used such assumptions in their empirical methodology. Grossman adds that there are other unobservable factors determining health, these include a child's natural abilities or qualities. These qualities are probably acquired at birth from the parents and he calls them *genetic endowments*. These factors can also be augmented by the mother's health status, mostly through antenatal nutrition.

In the model, it is assumed that a person inherits an initial stock of health and this health depreciates with time (considered as age). Grossman also assumes that this stock of health can be increased by a form of investment at one time in an individual's life. In this study we will work from a variation of the assessment model presented by Kutty (2008), which looks more closely at the influence of early childhood experiences on a child's long term health status.

The starting point of our Model is a household utility function given by:

$$U = u(H, C, L) \tag{4.1}$$

where utility depends on three factors: child health *H,* goods and services consumption *C* and the amount of leisure consumed *L*.

If we let T represent individual time inputs which includes allocated hours for child health/care T_h, for other income and non–income work T_w, and T_l is for leisure time, then households operate with a time constraint;

$$T = T_h + T_w + T_l \tag{4.2}$$

implying time available for work is $T - T_h - T_l$.

If we let P^h and P^c be the prices per unit for a child's health units E and goods and services consumption C respectively, the price for leisure consumed L be P^l, and that the market-determined wage rate be w, then we can express the household budget constraint as;

$$wT_w = EP^h + CP^c + LP^l \tag{4.3}$$

where EP^h, CP^c and LP^l are the health, goods and leisure expenditures respectively.

A further constraint is implied by the child health production function, which is given by;

$$H = h\left(E, T_h\right) \tag{4.4}$$

The mother's problem is to choose H, C, L to maximize U (equation (4.1)) subject to constraint (4.2) and (4.3). In this case we assume that $T > 0, T_w > 0, T_l > 0 \; and \; T_h$, and that

$$h_E > 0, h_{T_k} > 0 \tag{4.5a}$$

$$h_{EE} < 0, h_{T_k T_k} < 0 \tag{4.5b}$$

$$h_{ET_k} > 0, \text{ or } h_{ET_k} < 0 \text{ or } h_{ET_k} = 0 \tag{4.5c}$$

The subscripts denote the partial derivatives so first order derivatives of equation (4.5a) are the marginal utilities. As is standard in economic theory, these inequalities state that utility increases with higher levels of consumption of E and T_h. If a person has more time to work, that leads to more income and therefore a greater ability to purchase more health for the child. On the other hand there will be less time available to the child. Similarly, if a person has more time for their child that means they will have less time to work and hence have less income to purchase health units for their child. The inequalities (4.5b) state the law of diminishing marginal utility. This means that while utility increases with the consumption of E and T_h, it increases at a diminishing rate. The sign of h_{ET_h} in equation (4.5c) depended on whether E and T_h are substitutes or complements or unrelated. If E and T_h are substitutes, then we expect $h_{ET_h} < 0$; if complements, then $h_{ET_h} > 0$; and if at all are unrelated, then $h_{ET_h} = 0$.

Applying the method of Lagrange, we introduce the multiplier λ and form the Lagrangian

$$\begin{aligned} V = {} & u(H, C, L) + \lambda_1 (T - T_h - T_l) + \lambda_2 (wT_w - EP^h - LP^l) \\ & + \lambda (H - h(E, T_h)) \end{aligned} \tag{4.6}$$

where λ_j are the Lagrange multipliers giving the rate of change of the optimal value relative to constraints in each bracket. To maximize utility, the optimal quantities,

say H^*, C^*, L^* and the λ^* necessarily satisfy the first order conditions given by the partial derivatives of V in (4.6).

The child's health production function can now be obtained by working from the first order conditions derived from the above constrained maximization problem in equation (4.6). The child's health production function will be given in the form:

$$H = h\left(X_h, X_c, X_e, P, \varepsilon\right) \tag{4.7}$$

where X_h consists of health inputs to the child (biomedical factors); this includes nutritional intake, medical services such as immunization vaccines, and the time and care services the child receives from the mother and probably any other household member such as elder siblings or relatives; X_c is the factors consisting of the child's characteristics which include gender, size at birth, etc; X_e is residential factors (such as region where the household resides, environmental situation, etc), the stability of the household (eg. household dwellings), the mother's characteristics (can be seen as direct inputs) and such related factors which may have an effect on a child's health after birth; and ε being non–income household production factors and other unobservable factors for the child from the mother before birth (*in utero*), other household members, social status in the society or community and environmental factors which have effect on a child's health but cannot be measured.

The empirical modeling to follow will use the child health production function described in equation (4.7).

4.4 Econometric Methodology

In this section we give variable definitions and specify the empirical Model to be used in our analysis.

4.4.1 Dependent Variable

As a measure of child health, we use the z-score for body mass index (*bmi*) for age as our main indicator for this study. In addition, we will also model the z-score for height for age and z-score for weight for age.

The z-score classification system is widely recognized as the best system for analysis and presentation of anthropometric data for both individual level and population based assessments. The formula for z-score computation is

$$Z-score=\frac{\text{Observed value } - \text{ median value of the reference population}}{\text{Standard deviation value of reference population}} \quad (4.8)$$

There are several advantages to using z-score. First, it is easy to interpret the results because the scale is linear and standardised, so z-scores can be compared across measures (*bmi,* height, or weight) and age. In other words, z-scores have the same statistical relation to the distribution of the reference around the mean at all ages, which makes results comparable across age groups and indicators. Similarly, z-scores are sex-independent and thus permit the evaluation of children's growth status by combining sex and age groups.

To obtain the z-scores for these indicators; height for age, weight for age and *bmi* for age, we use an extension procedure (*zanthro*) for Standardizing anthropometric measures in children and adolescents in STATA (Suzanna *et al*., 2004).[50]

Note this STATA function uses US or UK standards. The US method calculates z-scores using 2000 *Centers for Disease Control and Prevention* (CDC) growth as the reference point, while the UK method uses the 1990 British Reference as the reference Data. The UK standard has been used in this study as it provides data for ages across the whole range. However, since our data is from a different geographical zone or continent, we were not sure if the use of this method would produce appropriately standardised z-scores for this Kenyan data.

To examine this issue we plotted histograms of the z-scores for two different ages of children; 0 – 18 months and 18 - 54 months, looking at all three health measures (refer to Figures 4a – 4c). From these graphs, we see that although the z-scores appear approximately normally distributed, the mean values are negative for *bmi* for age and even more negative for height for age and weight for age. They are also more spread than $N(0,1)$. This indicates that, on average, the values of these indicators ought to be standardised by values quite different from the 1990 British Reference Data. However, since our interest is to use these indicators to assess child health and not to produce health charts, then this itself poses no major problem for our analysis. However, if the z-scores were not distributed similarly across the ages of the child, then this would be a problem for our estimations and alternative measures would be required. The plots by

[50] For details see Kuczmarski *et al.*, (2002); Cole, *et al*., (2000); Cole, *et al*., (1998); Cole & Green. 1992 and Cole, (1990)

age groups indicate that our z-scores for each health indicator are distributed quite similarly across the age ranges.

4.4.2 Independent Variables

The structural model includes several explanatory variables. The first relevant variable is the child's age. Due to age's potential non–linear effect on child health, we estimated a piece-wise age spline with three knots at 6, 18 and 54 months.

A child's size at birth is most likely a key determinant of health. Birth size was classified into five ordinal classes: very large, larger than average, average, smaller than average and very small. A dummy variable for if the child is born as a twin was also included. The next set of child characteristics measure nutrient intake. These include the number of months the child was breastfed, the number of times the child was given water, juice, commercially produced baby milk formula, powdered or fresh animal milk, pumpkins and/or carrots or red/yellow yams or red sweet potatoes etc, green vegetables or leaves, fruits rich in vitamin A such as mangos, papaya, etc, food made from local grains. We also include dummy variables for child vaccinations. These included B.C.G (for tuberculosis), D.P.T (called triple vaccine: for Diphtheria, Pertussis Whooping Cough and Tetanus) and Polio.

A dummy variable for a child's sex was also included. Some literature has indicated that birth order can be an endogenous factor, especially in developing countries where the parents would like to have a male child in preference to a female child as a first born or at least to have a male child (children) in their family. They may end up having many children, not out of choice or intention, but as a result of attempting to have a male child

because they have had female children in the previous birth(s). This brings in an element of endogeneity for a child's gender to the mothers fertility (Horton, 1988). As a result they may end up having big numbers of children and due to the limitations of the household resources, this can affect the children's health due to a shortage of adequate health investment.

Apart from a child's characteristics, we include variables to capture the mother's characteristics and their influence on their children's health. These variables include a mother's educational attainment, which was constructed with three dummy variables: no education at all, at least primary school education and at least secondary or higher education level. If the mother had contracted HIV/AIDS, this was captured by a dummy variable. As noted earlier, the mother's HIV status is not observed for all cases, as participation in the HIV test was voluntary. We tried two approaches to deal with this issue: a reduced sample with only observed data (Model 1) or to impute the missing observations (Models 2 – 5). As we found in Chapter 3 on child schooling, the two methods gave similar results. We also include the variable for a mother's body mass index (*bmi*) and a dummy for the mother's experience of violence with her partner, taking the values one if violence has been experienced and zero if no violence was experienced.

There are also several household variables included in the model. They include the household size, a dummy variable for whether the child lives in an urban or rural area and a set of dummy variables for the eight provinces of Kenya. Another variable included in the model was household wealth. This variable was given in quintiles which were classified into poorest, poorer, middle, rich and richest. As in previous chapters, we argue that wealth is likely to be endogenous in this model. To deal with this, we

constructed a continuous variable from the household's assets using principle component analysis. This enabled use of the instrumented variable (IV) procedure by including the residuals from an equation explaining wealth in the child health equations (Models 3 – 5).

A potential selection bias exists in that the sample only includes children who are living. To deal with this, Heckman's selection model was used. Fortunately, the DHS survey includes information about children who were part of the sampled household, but had died. This allows us to estimate a selection equation for a child's survival and use this to correct for selection bias in the child health equation. Several variables for the selectivity model were obtained; these were thought to contribute to a child's survival around birth or early infancy. In addition to those in the child health equation, identifying variables reflected investments of the mother that would enhance the child's chances of survival in infancy. The investments, apart from the mothers' education levels, include the mother's use of contraceptives (hormonal and non-hormonal), whether the mother had a tetanus injection during her pregnancy, whether the mother had pre-natal clinic visits during pregnancy, a variable capturing if the mother took iron tablets during her pregnancy, a dummy variable for if the delivery took place at a hospital, a dummy for if the baby was born by caesarean, and finally, we included a dummy variable for whether the mother first had sexual intercourse under the age of 15 years. These variables were checked to see if they would have any impact child health and all were insignificant.

4.4.3 Empirical Specification

We defined the child health production Model in section 4.3.2 in equation (4.7), and in this section the estimation strategy will be discussed.

There are a number of difficult econometric issues to address in the estimation. Firstly, household wealth is likely to be an important explanatory variable, but there are issues with the measurement of wealth: in the DHS survey, there is no continuous wealth or income measure. Instead, each household is placed in one of five wealth categories, based on their status with respect to various assets. In addition, wealth is likely to be endogenous in the child health equation: there are unobservables that would affect both the household's wealth and the health of their children. A second issue surrounds the HIV variable. As noted in previous chapters, not all adults were tested for HIV status as part of this survey, so there are many observations missing on the dummy variable indicating the presence of HIV/AIDS among adults in the household. The third issue surrounds potential selection bias in examining health of children, in that the only children included in the sample are those who are still alive, having survived early childhood. It is very likely that there would be unobservables which would influence a child's survival (selection into the sample) and their current health status, producing a selection bias. We will next outline in some detail how we will deal with each of these econometric issues.

The starting point for the estimation is the following structural equation:

$$Z_i = X_i^{'}\beta_1 + W_i^{*'}\beta_2 + H_i^{'}\beta_3 + u_{1i} \qquad (4.9)$$

Equation (4.9) is the base Model we used to estimate the child's health production function, where Z_i is one of the z-scores of the measures of a child's health (height, weight, *bmi*), X_i is the exogenous set of covariates comprising the child, mother and household characteristics, W_i^* is the variable measuring household wealth and H_i is a dummy variable indicating whether the mother was tested as HIV positive.

$\beta = (\beta_1, \beta_2, \beta_3)'$ is the vector of parameters we want to estimate and u_1 is the error term such that $E(u_1) = 0$.

Firstly, there are significant issues around wealth W_i^*. As noted in earlier chapters, the DHS data set does not provide a continuous measure of wealth. Instead, based on responses to 22 questions around household assets, a principal components analysis was used to construct an index of wealth. This data set allocates each household to one of 5 wealth categories from poorest to richest depending on their value for the wealth index. So one option for capturing wealth effects on children's health is to include this set of dummy variables capturing the different wealth categories for the household. There are two problems with this. First, it discards information about wealth – a continuous wealth index has been replaced with a set of 5 categories. There is a great deal of variation within these categories that is ignored here. Secondly, wealth is almost certainly endogenous in equation (4.9), and dealing with an endogenous ordinal variable is not easy. For these reasons, we have followed the DHS process, but stopped one step from the end. Namely, we use principal components analysis on the asset indicator variables to construct, using the first principal component, a continuous index of household wealth.

Two methods were used in selecting the components to be used in the analysis. Firstly, we used the criteria which was developed by Kaiser (1960) called the Kaiser Criterion. With this approach, you retain and interpret any component with an eigenvalue greater than 1.00. In this criterion, each observation contributes one unit of variance to the total variance in the data set. Any component with an eigenvalue greater than 1.00 is accounting for a greater amount of variance than had been contributed by one

observation. In this case, such a component is therefore accounting for a meaningful amount of variance, and is worthy of being retained (Kaiser, 1960). Another method we use to confirm this component retention was the screeplot. In this method, the eigenvalues associated with each component are plotted. One is required to get the component just before a "break" between the components with relatively large eigenvalues and those with small eigenvalues. Again, the components that appear *before* the break are assumed to be meaningful and are retained for rotation whereas, those appearing *after* the break are assumed to be unimportant and are not retained (Cattell, 1966). From these two methods, only component one was retained as a continuous measure of wealth as a proxy for household wealth. Note that wealth is almost certainly measured with error, another reason to allow for potential endogeneity of this wealth variable. This endogeneity is dealt with using standard instrumental variables, with asset variables and other household characteristics (eg. Education level of household head) as instruments.

Specifically, we specify the following equation:

$$W_i^* = X_{Wi}^{'}\gamma + u_{2i} \tag{4.10}$$

Endogeneity of W_i^* in equation (4.9) is represented by the assumption that $Cov(u_1, u_2) \neq 0$. X_{Wi} is a set of variables that are correlated with household wealth, including some variables not in X_i or H_i in equation (4.9) – these are the instruments for W_i^* which are uncorrelated with u_{1i}. To correct for bias due to this wealth endogeneity, we obtain estimated residuals from an estimation of equation (4.10) and use these together with W_i^* in our main equation (4.9).

The second problem in estimating equation (4.9) is with H_i, the variable capturing the HIV status of the mother. As discussed in chapter 2, there are many missing data points for this variable, as by design, HIV testing was conducted on only around 50% of adults in the sample. This means a sizeable loss of sample size in our estimation of (4.9). To avoid this problem, we used the method of imputation by chain equation (ICE) to impute the missing data. To do this we specify:

$$H_i^* = X_{Hi}^{'}\delta + u_{3i} \tag{4.11a}$$

$$H_i = \begin{cases} 1 \; if \;\; H_i^* > 0 \\ 0 \;\; if \;\; H_i^* < 0 \end{cases} \tag{4.11b}$$

$$\widehat{H_i} = \begin{cases} 1 \; if \;\; X_{Hi}^{'}\widehat{\delta} > 0 \\ 0 \;\; if \;\; X_{Hi}^{'}\widehat{\delta} < 0 \end{cases} \tag{4.11c}$$

$$\widetilde{H_i} = \begin{cases} H_i \; if \;\; H_i \; is \; observed \\ \widehat{H_i} \;\; otherwise \end{cases} \tag{4.11d}$$

H_i^* is the latent variable and H_i is our observed variable. Equation (4.11a) describes the equation by which the relationship between HIV status and observable variables is estimated for those adults where we have data on HIV status. X_{Hi} is observed for all (or at least most) households, and also includes some variables not in equation (4.9). Based on estimates from (4.11a), the HIV propensity is estimated for other adults who were not tested, and an imputed value of HIV status obtained.

The other problem to deal with in our estimation of (4.9) is the potential for non-random selection due to the fact that only children who have survived to the time of data collection are actually included in the sample. There are very likely to be unobservable factors that contribute to survival through early childhood, and which also influence the current health status of a child who has survived. This is a standard problem of sample

selection based on unobservables, and can be dealt with using the commonly applied Heckman procedure where we specify:

$$S_i^* = X_{Si}^{'}\psi + u_{4i},\ and\ S_i = \begin{cases} 1\ if\ \ S_i^* > 0 \\ 0\ otherwise, \end{cases} \qquad i = 1,2,...,n \tag{4.12}$$

where S_i^* is the latent variable capturing propensity to survive, and S_i is the observed child's survival. Potential bias in selection is captured by assuming $Cov(u_{4i}, u_{1i}) = \sigma_{u_4 u_1} \neq 0$. Equation (4.12) can be estimated because the DHS survey collects information from each household about children who have died (stillborn or died afterwards), as well as those who are still alive. The covariates in equation (4.12), X_{si}, include factors which would influence the child's survival, and for identification, at least one variable in X_{si} needs to not belong in equation (4.9) – in other words, they are factors that influence a child's early survival, but not their current health status except via their impact on survival. As noted earlier, we choose variables that specifically relate to the birth experience of children, as these are likely to have their effect almost exclusively on survival, and not on current health status.

Following the standard Heckman procedure (see Wooldridge, 2006, p. 618-620), since the conditional expected value of Z_i given $S_i = 1$ is

$$E(Z_i \mid S_i = 1) = X_i^{'}\beta_1 + W_i^{'}\beta_2 + H_i^{'}\beta_3 + \sigma_{u_4 u_1}\lambda\left(X_{Si}^{'}\psi\right), \tag{4.13}$$

and $$\lambda(\alpha) = \phi(\alpha)/\Phi(\alpha)$$

where α is a real number over $(-\infty, \infty)$, $\phi(\alpha)$ and $\Phi(\alpha)$ are the *pdf* and *cdf* of the standard normal distributions respectively, equation (4.9) can be re-written as

$$Z_i = X_i^{'}\beta_1 + W_i^{*}\beta_2 + H_i^{'}\beta_3 + \sigma_{u_4 u_1}\lambda\left(X_{Si}^{'}\psi\right) + \in_i \qquad (4.14)$$

An estimation of the parameters of equation (4.12) can be used to estimate the inverse Mills ratio, $\lambda\left(X_{Si}^{'}\psi\right)$, and this would then be included as an extra variable in the estimation of equation (4.9), as shown in (4.14).

To summarise, we have a linear equation for Z_i, the child's current health status. To estimate over the maximum possible sample, we estimate a separate Model for adult HIV status which can be used to impute the HIV status of mothers in the sample for which an HIV test was not conducted. The child health equation is then estimated using instrumental variables to deal with endogeneity of wealth (via inclusion of residuals from the wealth reduced form equation) and allowing for sample selection due to some children dying (by addition of an inverse Mills ratio obtained from the estimation of an equation explaining a child's survival).

This set of equations could in principle be estimated as a system using maximum likelihood, but given the complexity of the system, with four equations to be estimated and nonlinearities present in different forms, it is likely that we would experience severe convergence problems. Integration of the likelihood function across four dimensions is computationally intensive, and often unreliable. Consequently we have opted for the slightly more ad hoc, but hopefully more reliable, approach of augmenting the main equation with the appropriate residuals to give consistent estimates. While there is a small loss in efficiency compared with maximum likelihood, the estimation is much less vulnerable to convergence and stability problems.

4.5 Discussion of Results

In this chapter, we are using the same data set used in previous chapters, the Kenya Demographic Health Survey, 2003 which is the latest in a series of national level population and health surveys carried out over the last three decades in Kenya.

4.5.1 Descriptive Statistics for Selected Variables

In this chapter, for the data analysis, we used STATA/IC 11 for Windows, (2010). Table 4.2 gives descriptive statistics for some selected variables used in our Models. The table indicates that there are 5,949 children aged up to 59 months. Among them, about 25.8% came from urban areas and 74.2% from rural areas. In total, there were about 49% females and 51% male children. Out of these children, about 91.6% of the children were alive and 8.4% were deceased. The selected variables used included weight and height. 4,958 were measured for weight with a mean value of 11kg and 4,887 were measured for height with an average height of 82cm. The other variable was the child's size at birth: very large children comprise about 5.2% of sample; with 6.9% from urban and 4.6% from the rural areas. Those larger than average in body size were about 19.1% of the sample with 16.9% from the urban areas while 19.9% came from the rural, 16.8% females and 21.4% male children. 58% of children were born with average body sizes, 12.3% with smaller than average body size and those who were very small were 4.3% of the sample, but with 5.4% females and 3.3% male children. We found in total, 3.4% were born twins.

The variables for nutritional intake capture the effect of diet or nutrients given to the child as they grew up. To mention a few, on average the number of months a child was

breastfed was 2.8, with an average of 2.6 months for the urban and 2.8 months for those from rural areas. There were huge difference in the proportions of the children given most of the foods in the urban versus rural areas. Out of 4,687 children, 84.3% of urban children were given water on a frequent basis compared to only 77.5% from rural areas. Only 19.1% of the overall sample were given juice, with a much larger proportion of 32.4% for the urban, while in rural areas only 14.5% were able to be given juice. The same trend shows up with children given commercially produced baby formula, children given pumpkin, carrots, red or yellow yams and red sweet potatoes and fruits rich in vitamin A such as mango, papaya etc. However, the children who were given powdered or fresh animal milk, green leafy vegetables and food made from local grains had more similar proportions in the urban and rural areas. In our estimations, these variables including the months of breastfeeding could lead to a censoring issue in that at the time of survey, some children were still being breastfed as well as being given food. However, this problem is controlled for by including the child's age in the models.

In the sample, 84.1% were vaccinated against *bcg* across the country, with a higher vaccination rate in urban areas, but no real gender differences. Other vaccinations followed the same pattern, with rates of 83.5% against *dpt* (Diphtheria, Pertussis Whooping Cough and Tetanus diseases) and 85.1% against polio.

We now turn to variables describing the mothers' characteristics. The proportions of mothers' education levels showed that 20.3% of mothers had no education, 58.1% had at least primary education and about 21.6% had attained at least secondary or higher education level. As expected, the rates for mothers without any education were higher in the rural areas with the value of 22.4% compared to the urban areas with 14.3%. In contrast, for the mothers with at least secondary or higher education level, the proportion

was higher in the urban areas with 36.8% compared to 16.3% from rural areas. On average, 8.2% of the mothers were HIV/AIDS positive with a higher incidence of HIV in urban areas (12.3%) compared to the rural areas with just 6.9%.

To construct the household wealth variable, we obtained the first component from the Principal Component analysis on a range of household asset indicator variables. As we indicated earlier, we calculated a KMO value for household assets, which was 0.969. KMO is a measure of the sampling adequacy used for comparison of the magnitude of the observed correlation coefficients and the magnitudes of the partial correlation coefficients. This value was very far from a cut off point of 0.6 which means it is valid to use Principal Components in these household asset variables. We used Bartlett's test of sphericity to test the null hypothesis that the assets correlation matrix was uncorrelated. We were able to reject the null because of high value for the computed chi-square, $\chi^2 = 2.32e+05$ and with 66 degrees of freedom, the $p-value = 0.000$, indicating that the relationship among these household assets was very strong. The scree plot revealed that we ought to pick one component since the elbow occurred at component 2 (Figure 4.1d). This suggests that one component accounts for a very large share of the combined variance. The eigenvalue for this component was 9.730 and for component 2 was 0.925 which also indicates that it was enough to retain the first component.

4.5.2 Models

We report various estimation results for a range of models, along the lines outlined in section 4.4.3. Dependent variables are various measures of a child's health: z-score for the child's body mass index (*bmi*) for age, z-score for the child's height for age and z-

score for the child's weight for age. The different model specifications are summarised in the table below.

Table 4.1. **Summary of Equations used in the Models**

Model	Equation used (refer to section 4.4.3)	Explanation
1	4.9 and 4.13b	• Does not deal with endogeneity of wealth • Uses only observed data for maternal HIV/AIDS status • Does not deal with child's survival selectivity bias
2	4.9, 4.11d and 4.12	• Does not deal with endogeneity of wealth • Uses observed and imputed data for maternal HIV/AIDS status • A Heckman selectivity Model was fitted for child survival
3	4.9, 4.10,4.11d and 4.12	• Uses Instrumental variables to correct for wealth endogeneity • Uses observed and imputed data for maternal HIV/AIDS status • A Heckman selectivity Model was fitted for child survival
4 & 5	4.9, 4.10, 4.11d and 4.12	As per Model 3, but with separate estimates for female and male children

The set of results associated with Model 1 can be thought of as the base model, where we estimate child health equations by OLS with just the available data, ignoring potential endogeneity and selectivity bias. Model 2 seeks to significantly increase the sample size by imputing the HIV status of mothers who were not tested, and to deal with potential selectivity bias due to early childhood mortality. Model 3 then adds to this scenario an adjustment for the endogeneity of wealth, thus representing the "complete" Model with all the obvious econometric issues addressed. Models 4 and 5 then provide

separate estimates when the sample is split into males and females, given the likelihood of a number of potentially different gender effects.

The results are given in Tables 4.3, 4.4 and 4.5. The associated selectivity Model results are given in the same Tables, after the Model estimates. Table 4.3 provides estimates where the child's health is measured by *bmi*, Table 4.4 uses height and Table 4.5 uses weight as an indicator.

4.5.2.1 Z-score for bmi for age

From Model 1 (Table 4.3), the gender dummy is not significant, indicating there is no difference in z-score for *bmi*-for-age and gender of the child. As noted earlier, we estimated the effect of age with a spline using three knots; at 6, 18 and 54 months. To interpret the effect of a child's age we plotted the net coefficient values of the three age splines[51] against child's age (Figure 4.2a). Firstly, looking to the age coefficient before spline transformation, we find it very significant with a negative coefficient of -0.278 ($p<0.01$) meaning the *bmi*–for-age z-score decreases steadily at a rate of about 0.278 standard deviations for every one month increment of a child's age. Recall this result is obtained relative to the UK reference data indicating the distribution of Kenyan child's z-score for *bmi* for age is slightly on the left to that of UK child's z-score for *bmi* for age and the gap increases with age initially. Looking into piece–wise spline age of the child, we find that between the age of 0 – 6 months, *bmi* for age z-score decreases with a value of 0.278 standard deviations monthly, relative to the UK z-score standards. However, after 6 months, the *bmi* for age starts to increase at a value of 0.030 standard deviations till the age of 18 months. From 18 months to 54 months, the *bmi* for age z-score seems to

[51] $\beta_2 \text{child age}_i + \beta_3 \text{age above6}_i + \beta_4 \text{age above18}_i + \beta_5 \text{age above54}_i$

be constant, with a value of about -0.003 standard deviations. However, after 54 months, the coefficient is insignificant although indicates a decrease with a rate of 0.047 standard deviations. This means that during the tender age, the child's *bmi* is likely to reduce relative to the UK but as the child grows up, probably from six months to one and half years, the *bmi* begins to improve significantly before it stabilises onwards. This is a clear indication that the *bmi* of children is most different to that for UK children in the first 6 months.

Apart from the child's age, a number of other child characteristics affect child health as shown in Table 4.3. These include the child's size at birth. This variable was categorised into five levels, where we find a child born average, smaller than average and very small in size have significant but negative coefficients of -0.370, -0.587 and -0.579 ($p<0.01$) respectively. This means that the child born with these body size levels is likely to decrease in *bmi* z-score by 0.370, 0.587 and 0.579 standard deviations respectively as they grow up compared to a child born with very large size. Looking into the child's nutritional intake, we find the variable for giving child green leafy vegetables to have a positive coefficient of 0.230 ($p<0.05$), meaning giving the child green leafy vegetable is more likely to give the child a *bmi* boost of 0.23 standard deviations as they grow. However, not all nutritional food gives improved health. We found the variables for giving the child commercially produced baby formula and giving the baby animal milk (either powdered or fresh) have negative coefficients of -0.363 ($p<0.1$) and -0.179 ($p<0.05$) respectively. These imply that a child given these types of milk will have a significantly lower *bmi*.

Turning to the household characteristics, four province dummies had negative and statistically significant coefficients. The provinces are the Coast, Western, Eastern, the

Rift Valley and North Eastern with coefficients of -0.314, -0.290 ($p<0.1$), -0.359 ($p<0.05$), -0.480 and -0.829 ($p<0.01$) respectively.

One important household variable included in our model was wealth, measuring the household wealth effect on a child's *bmi* for age. We found that coefficients of all wealth dummy variables were insignificant. This suggests that household wealth does not affect the child's health, specifically, the child's *bmi* for age.

There are characteristics of the mother which affect child health. We first consider mother's education level. Compared to a mother without any primary school education, the one with at least primary school education and with at least secondary or higher education had significant positive coefficients of 0.265 and 0.431 ($p<0.01$) respectively. This indicates that a mother with at least primary school or higher education level is more likely to improve the child's *bmi* for age by 0.265 and 0.431 standard deviations. The other maternal characteristic which seems to improve child health in terms of *bmi* is the mothers *bmi* itself which has a coefficient of 0.00054 ($p<0.01$). This indicates a child with a healthier mother is more likely to be healthy too, where the mother's *bmi* increases by 1 kg/m^2 will increase the child's *bmi* for age by 0.00054 standard deviations.

The HIV variables (one with missing observations and the other with imputed missing data) were also included in our analysis. The variable with missing data was used in Model 1 while the one with imputed data was used in the other four models. From Model 1, this variable had a positive but insignificant coefficient revealing it appears to be irrelevant to the child's *bmi* for age.

Turning to the other models, Model 2 shows similar results to those in Model 1 apart from the variables for nutritional intake which are now insignificant. Also, in Model 2, we find insignificant coefficients for the Eastern and Western provinces. The variables for the household wealth and imputed HIV again have insignificant coefficients.

In Model 3, we incorporate both HIV imputed variable and household wealth with endogeneity bias correction. We also use the child survival selectivity Model to correct for possible bias due to sample selection. This model has similar results to that in Model 2.

Models 2 - 5 include an Inverse Mills ratio (IMR) to control for the unobservables that are correlated with the selection of children who survived into the estimation. Using Heckman's two step estimates, these Models gave a significant inverse Mills ratio of 0.512 ($p<0.1$) and 0.670 ($p<0.05$) respectively, indicating that ignoring this sample selection would lead to biased estimates. In addition, we estimated this Model again using maximum likelihood estimates and used the Likelihood Ratio (LR) test to test for the presence of unobservables that are correlated between the selection and the main equation; if the correlation ρ is zero, the selection equation is not needed. The result shows that for Model 2, $\rho = 0.328$ and is significant with $\chi_1^2 = 3.900$ $(p = 0.048)$ and for Model 3, $\rho = 0.370$ which is significant with $\chi_1^2 = 5.680$ $(p = 0.017)$ which confirms that there would be bias in the estimates if we ignored the sample selection issue.

We now discuss results for the model of child survival, used to model selection into the sample. The instruments included in this equation include certain characteristics of the

mother. The variable indicating if the mother had her first sexual intercourse before the age of 15 years old had a significant coefficient of -0.124 ($p<0.1$). This implies the child whose mother had sexual intercourse before the age of 15 years was less likely to survive. Regarding contraceptives the mothers had used, the dummy for hormonal contraceptive had a coefficient of 0.159 ($p<0.05$). This means that the mother using this type of contraceptive will likely increase the chance of her child's survival. Similarly, the variables for mother having tetanus injections during her pregnancy period and pre-natal clinic visits had a significant coefficients of 0.246 ($p<0.1$) and 0.360 ($p<0.05$). The coefficient for mother's *bmi* suggests that as the mother's *bmi* increases the less likely the child is to survive. This could contribute to high risks and difficulties during deliveries where most stillbirths occur. The tribal variables indicate if the mother is a Luo by tribe, the child is less likely to survive compared to a child whose mother is from other tribes.

The variable for household size has a coefficient of 0.126 ($p<0.01$) indicating the bigger the household, the more likely the child is to survive. This is not what is expected because an extra member in the household means they take an extra share of the household resources, which would add stress to poor families. To check if this reflected our data we cross tabulated the number of household members and the variable for child survival (Table 4.9). This table confirms this model's estimate in that for every extra household member the child survival rate also increases. The variable for a mother with HIV and violence with husband or partner had significant and negative coefficients of -0.185 ($p<0.1$) and -0.146 ($p<0.05$) respectively. These indicated that a child whose mother is HIV positive or experiences violence from her husband or partner is less likely to survive.

Turning to the other two models, Model 4 and 5 use the same model set up as Model 3, but analyse females and males separately. These two Models show similar results to Model 3 apart from a few cases. In Model 4, the variable for a child born as a twin has a significant and negative coefficient of -0.322 ($p<0.1$) meaning that a female twin is likely to have a lower *bmi* for age by 0.322 standard deviations relative to a child who is not a twin. The variable for a female child receiving the *bcg* vaccination has a significant but negative coefficient of -0.227 ($p<0.1$).

In Model 5 for males, we find the variable for a child's birth size mostly insignificant, only a child born with smaller than average size is significant with a coefficient of -0.326 ($p<0.1$). In addition, we find the variable for nutrient intake of green leafy vegetables has a positive and very significant coefficient of 0.269 ($p<0.01$). This implies that by giving your male child green vegetables the child's *bmi* is more likely to be higher. However, giving the child powdered or fresh animal milk will decrease the child's *bmi* for age by 0.149 standard deviations.

The selection models in both Models 4 and 5 have insignificant IMR as in Model 3. However, in Model 4, the Likelihood Ratio test had $\rho = 0.500$ and is significant with $\chi_1^2 = 7.88 \ \ (p = 0.005)$, implying the presence of unobservables in the selection model that are correlated with the main equation for the child's z-score for *bmi*-for-age. Model 5 had an insignificant test statistic for ρ.

Comparing the variables in the child survival equations, we find a few differences compared to Model 3. Both models have insignificant coefficients for mothers' education levels apart from the variable for a mother with at least primary education in

Model 4 with a coefficient of 0.364 ($p<0.05$). This indicates mothers' education does not improve the survival of a male child but mothers with at least primary education do have better female child survival. The variable for HIV is insignificant in both models, meaning the mother being HIV positive does not affect her child's survival. Model 4 has insignificant coefficients for the variables measuring the age at which the mother first had sexual intercourse, use of hormonal contraceptives, pre-natal clinic visits during pregnancy and being a Somali by tribe. These estimates appear to play no important role in female child's survival. In Model 5, we find the variables for a mother being a Mijikenda or Swahili by tribe do appear decrease the chance of male child's survival.

4.5.2.2 Z-score for height for age and weight for age

As alternative health indicators, we also used height for age and weight-for-age to analyse the child's health status. Again we used the z-scores (standard deviation scores) for these indicators which were computed using UK standards. In this case, we give results for only Model 3 – 5 for both indicators, with some comparison to the results of the z-score for *bmi* for age in the previous section.

4.5.2.2.1 Height for age

The estimates for the z-score for height for age are given in Table 4.5. We will focus on Model 3 (*column 3*). The explanatory variables are the same as in the previous section. In this case, we find the variable for a child sex is significant with a negative coefficient of -0.169 ($p<0.01$). This indicates that boys deviate from the UK standard height for age measure more than girls. Figure 4.2b shows the net coefficient values for the overall age and piece-wise splined age variables. The values show a huge difference from the *bmi*

health indicator. The age before transformation has weak significant coefficient of -0.056 ($p<0.1$). This means the amount by which the Kenyan child's height falls short of the standard increase by 0.056 standard deviations per month. Unlike the case of *bmi*, the gap remains constant for ages from 6 to 18 months, and becomes smaller in the age range of 18 to 54 months, the child's height gap closing by about 0.022 standard deviations compared to the UK height standard. As with the case of *bmi*, as the child grows older there is a smaller height difference between the Kenyan and the UK child.

Turning to other personal characteristics affecting the child' s own health, we have the child's body size at birth which gives similar results to the *bmi* for age health indicator. The variable for a child born as a twin is now significant with a negative coefficient of -0.698 ($p<0.01$) which means a child born as a twin is likely to be shorter by 0.698 standard deviations compared to the non-twin. The variable for a child being vaccinated against *dpt* has a coefficient of 0.252 ($p<0.05$) which means a child vaccinated with *dpt* (Diphtheria, Pertussis Whooping Cough and Tetanus diseases) is more likely to improve his or her height for age by 0.252 standard deviations compared to the child who is not given such a vaccination.

The variable for the child having been breastfed has a significant but negative coefficient of -0.131 ($p<0.01$) compared to an insignificant coefficient in the case of *bmi*. This is unexpected as it means breastfeeding a child will reduce the height of the child. In contrast, unlike the case of *bmi*, we get a positive effect for the variable of the child being given commercially produced baby milk formula which has a coefficient of 0.456 ($p<0.01$). This imply that giving a child commercially produced baby milk formula is likely to improve the child's height for age.

There are a few household variables which are significant when compared to *bmi* for age, outlined in the previous section. These include the household size with a coefficient of -0.027 ($p<0.01$) indicating that for every extra household member the height of the child is likely to reduce by 0.027 standard deviations. The other household variables affecting the child's height for age and not *bmi* for age are the dummies for Nyanza and North Eastern provinces. Both have significant positive coefficients telling us that a child from these provinces is significantly taller than the base province, but does not have a different *bmi*. In this model, the household wealth variable is now significant with a positive coefficient of 0.771 ($p<0.01$) unlike the case of *bmi*. This indicates a strong positive effect of wealth on the child's height meaning that every unit increment of wealth the child is more likely to increase in height by 0.771 standard deviations. The same effect is observed in Models 1 and 2. In Model 1, the variables for the poorest and poorer households have negative coefficients of -0.540 ($p<0.01$) and -0.297 ($p<0.1$) for height-for-age respectively. For Model 2 the four variables for households being poorest, poorer, medium and richer have significant and negative coefficients of -0.611, -0.338, -0.245 and -0.232 respectively, all quite strong negative effects compared to the child from the richest household. These results are in line with what would be expected.

The variables for mothers' characteristics give similar results as in *bmi* for age cases.

To give a brief overview of the selection Model for Model 3, we find a significant inverse Mills ratio of -1.153 ($p<0.01$) unlike in the case of child's *bmi* for age. In addition, we used Heckman's likelihood ratio estimation, and the LR test for the selection equation (testing if H_0: $\rho = 0$) found $\rho = -0.232$ was significant with $\chi_1^2 = 4.440$ ($p = 0.035$). This implies significant selection on unobservables. The

variables with significant coefficients in the selection equation (Table 4.6) are the same as in the case of *bmi*-for-age, apart from the variables for mothers with HIV which has an insignificant coefficient.

Models 4 and 5, in Table 4.5, give the results for samples with just female and male children respectively. The estimates in Model 4 are based on 1,892 observations (female children) and Model 5 is based on 1,944 male children. There are a few differences in the estimates in Model 4 as compared to Model 3, but Model 5 has generally similar results to Model 3. In Model 4, the variable for an overall child's age is insignificant but the splined age at 18 months is significant with a coefficient of 0.123 ($p<0.01$). The child's birth size and the child being born as a twin has similar results in Models 4 as in Model 3. The other variables with different results from Model 3 are for the vaccines and nutrient intake.

The female child's height for age is more likely to be boosted by giving the child *dpt* vaccine with a difference of 0.323 standard deviations for age. The same effect is found in the variable for giving the child green leafy vegetables where the child is more likely to increase her height for age by 0.212 standard deviations. The strongest effect is for Nyanza province where the difference in height for age is 0.495 standard deviations for age compared to a female child from Nairobi province. The other variable is the household wealth which has a significant coefficient of 1.255 ($p<0.01$) indicating that a female child's height for age is more likely to increase by 1.255 standard deviations for age as household wealth increases by one unit.

Turning to Model 5, the main difference compared to the model for *bmi* for age is with the variable for the polio vaccine with a significant coefficient of 0.343 ($p<0.01$). A

similar effect is observed on the variable for food made with local grains which is more likely to boost the child's height by 0.211 standard deviations for age. It is worth mentioning that, unlike the female child, household wealth has no significant effect on a male child's height for age.

The inverse Mills ratios for Models 4 and 5 have significant coefficients at the 5% and 10% levels of significance respectively. However, the LR test for correlation in the errors gives $\chi_1^2 = 2.380$ ($p = 0.123$) and $\chi_1^2 = 1.560$ ($p = 0.212$) respectively and therefore, we could not reject null hypothesis that there is no correlation between errors of the selection and the height equation.

4.5.2.2.2 Weight for age

The results for the z-score of weight-for-age are given in Table 4.7 *column 3*, based on the specification of Model 3 with a sample of 3,947 children. This Model has different results compared to the results for *bmi*-for-age and height for age child health indicators. The variable for the child's sex gives the same results as for the child's height for age and similar to that for *bmi*-for-age. The variables for child's age show significant differences to the reference group, a child in UK. As in the case of *bmi*, the child's age before transformation has a coefficient of -0.231 ($p<0.01$). This means that overall the distribution of Kenyan child's z-score for weight-for-age is below that for the UK distribution and the gap widens with age. As we included piece-wise spline age in the model, the coefficients for the age of 6 – 18 months and 18 – 54 months had opposite signs to the *bmi* case, with the values of -0.059 and 0.010 ($p<0.01$) respectively. These coefficients imply that at the age 6 – 18 months, a Kenyan child's weight will continue to fall further short of the UK norms by an average of 0.059 standard deviations. As the

child grows up, at the age between 18 to 54 months gap is reduced by 0.01 standard deviations (See Figure. 4.2c).

When we look into other children's characteristics, they give the same results as for height for age except for the variables for household size and violence with husband or partner.

Compared to the *bmi* health indicator, we find a few differences that are worth mentioning. The variable for child born as a twin has a coefficient of -0.456 ($p<0.01$) implying a twin child is likely to weigh about 0.456 standard deviations less compared to a non twin child. The variable for a child vaccination against *dpt* has a coefficient of 0.273 ($p<0.01$) meaning vaccinating a child against *dpt* is more likely to improve the child's weight by about 0.273 standard deviations compared to the child who is not given such a vaccination. Contrary to what we expect, the variable for breastfeeding a child has a significant negative coefficient of -0.096 ($p<0.01$) compared to an insignificant coefficient in the case of *bmi*. This means breastfeeding a child will reduce the relative weight of this child.

In the case of household variables, we get a few significant values compared to the *bmi* health indicator. The household size variable has a coefficient of -0.025 ($p<0.05$) implying for every extra household member in the same household, the weight of the child in this household is likely to reduce by 0.025 standard deviations. The other household variable is the dummy for Nyanza province with a coefficient value of 0.258 ($p<0.05$). This tells us that a child from the province will be heavier by about 0.258 standard deviations than the base province. No difference was found for the case of *bmi*. The other variable we find with significant effect is the continuous household wealth

variable, with a coefficient of 0.671 ($p<0.01$). This means that for every unit increment of household wealth, a child is more likely to increase weight by 0.671 standard deviations. The last variable with an effect on a child's weight is violence with husband or partner, which has negative coefficient of -0.115 ($p<0.05$). This implies that violence between the mother and the husband or partner will reduce a child's weight by about 0.115 standard deviations compared to a child with no violence between the parents. These two variables, household wealth and the dummy for violence, had insignificant coefficient values in the case of *bmi*.

The selection Model for Model 3 gives an insignificant inverse Mills ratio. This is further confirmed by the Model's estimates using Heckman maximum likelihood estimates, where the LR test (testing if H_0: $\rho = 0$) with the $\rho = -0.317$ is insignificant with $\chi_1^2 = 2.550(p = 0.110)$. Thus, the unobservables in the selection model appear to be uncorrelated with the main equation of child's z-score for weight-for-age.

The results for Models 4 and 5 are based on 1,949 observations of female children and for Model 5 are based on 1,998 male children (Table 4.5 *columns 4* and *5*). There are no major differences in these results as compared with to the results from other cases of child's *bmi*-for-age and height for age.

4.5.3 Discussion

The above results are based on different models, with different estimation methodologies and measures of the child's health status. For the three child health indicators or measures, child health has been modelled taking into account some of the modelling issues, namely endogeneity and sample selection. In this section, we will

compare and discuss the above results. First, we will look into the structural equation estimates, then sample selection model estimates and lastly the study's contributions.

A child's health is affected by several factors during their growth, especially when they are under five years. Although there are other factors to be looked at, one of the most critical factors when the child is between 0 and 5 years of age are the characteristics of the child's mother. At this age the child relies on their parents or guardian who is in a position to invest in the child's health.

The results from the three sets of analysis show that there are some overall gender differences. Using height for age and weight for age, boys have inferior z-scores than girls, although *bmi* z-scores are no different. This implies that height for age and weight for age of males are more inferior to UK standards than females but that *bmi* are not. The reason for the inferior height and weight measures could not be established in this study, but it is probably due to biomedical or/and environmental factors. The other possible explanation would be due to immunological endowment from the mother. Our findings are in line with what has been found by earlier studies based on the Demographic Health Surveys. In their study, Hill and Upchurch (1995) using the data for Demographic and Health Surveys found that female child mortality rates were lower than the male children. However, when allowances for standard historical levels were included, a pervasive pattern of female disadvantage emerged. This was found to be true for a few countries, whereas Kenya still showed female children having an advantage over male children. Based on their study, they pointed out that it was important that further study be carried out to understand this gender difference in child mortality. They further suggested other factors surrounding socio-cultural and health issues that need to

be explored. Our study has extensively included such factors but still we were unable to establish fully the cause of this difference.

The results reveal that child health measures relative to the UK standard vary with the age of the child. The three indicators showed that during the first 6 months, the child's *bmi*, height and weight are inferior to the standard child in the UK, and the gap between the two groups of children widens with age. However, after six months, the gap for *bmi* starts to reduce and stabilises after about eighteen months. This was different for the case of height and weight. For the age between six and eighteen months, the gaps further increase and thereafter narrow compared to the UK standards. It appears that Kenyan children start with significant disadvantage compared to those from a developed country, and that in the early months, this disadvantage worsens, before some improvement is experienced. This suggests generations of economic disadvantage in Kenya have its greatest effect on child health in the first year of their life, consistent with vast differences in infant mortality rates. This is consistent with results in the study carried by Ruger and Kim (2006) where for 23 countries they found high child mortality in western and sub-Saharan Africa and Afghanistan. These countries also had significantly higher rates of extreme poverty.[52]

A child's size at birth is another important factor we found to affect their health. All models in the three analyses showed that, compared to the child born with very large size, a child born average, smaller than average and very small in size are likely to have lower *bmi* for age as well as height for age and weight for age. This is expected and makes sense. If a child is born small in size it is likely that their weight and height will be relatively poor due to an underdeveloped immune system at birth. The reason behind

[52] See also Day, *et al.* (2008); Van de Poel, *et al.* (2008).

the child's body size could be related to biomedical factors, especially from the parents. Our study reveals similar results to earlier research. Based on the study by Law and his colleagues on child's body size at birth and blood pressure among children in developing countries, it was found that blood pressure was positively related to the child's current weight. After making adjustments for weight and gender, systolic pressure was inversely related to size at birth in almost all countries they studied - Chile, China and Guatemala, although the results were not as clear cut for Nigeria (Law *et al.*, 2001. See also Forssas *et al*., 1999).

Considering nutrient intake, the three indicators showed some conflicting results. Using child *bmi*-for-age, Model 5 showed the male child increased his *bmi* by 0.269 standard deviations when fed with green leafy vegetables. This is expected as green vegetables are a very good source of Vitamins (C, A, E etc), minerals such as Calcium, Magnesium, Iron, etc and as well as a good source of Phytochemicals (eg. Beta-carotene) and Antioxidants, hence a good source of fibre. All these are essential in the body and vital as the child grows (Faber *et al.*, 2007). The effect of giving the child the green vegetables as part of their nutritional intake, agrees with the findings by Sanghvi and Murray (1997) on their study on improving child health through nutrition. However, using a child's height for age, giving the male child the same green vegetables reduced the height for age of the child by 0.295 standard deviations for age. It appears for boys, the regular consumption of vegetables is associated with shorter children with healthier body mass.

Breastfeeding and the use of commercially produced baby formula is a controversial issue. Results suggest the length of time child is breastfed and use of formula or animal's milk instead of breastfed milk has no impact on *bmi*. Further, the longer a child

is breastfed, the smaller both their height for age and weight for age. However, these cancel out to give the neutral effect on *bmi*. The results also reveal some evidence of a positive effect of formula and animal's milk on height for age.

Despite mother's milk having neutral effect on *bmi*, it gives a child immunological benefits **(**Field, 2005). It has been mentioned in previous work that a shorter duration of breastfeeding could be a predictor of adverse mental health outcomes (See Oddy *et al.*, 2010). The study by Zhou *et al.* (2005) on the impact of parasitic infections and dietary intake on children's growth in China found that the children who were retarded in growth when compared with the standards of Chinese rural children (especially the girls who were more frequently infected by *S. japonicum)* had lower intakes of protein and energy. Their study found that reduced height for age, weight for age, and mid-upper arm circumference were observed in children infected with *S. japonicum*, most severely in girls with the least energy and protein intakes.

In our study, breastfeeding appears not to help *bmi*. This could be explained by several issues which could not be addressed by our data. Firstly, Kenya is an multi-ethnic nation with more than 70 ethnic tribes who are grouped into three different groups namely Bantu (agriculturalists), Cushite (nomadic pastoralists) and Nilotic (pastoralist and fishermen). These tribes have different beliefs, cultures practices and traditional customs including different food styles which causes them to differ in their weights, heights and *bmi.* The other issues which could possibly explain breastfeeding not having significant effect on child's *bmi* could be the selection issue. It is more likely that people from the poor rural areas do not use the baby formula milk compared to those in urban areas. As such a child in the rural area is purely breastfed while the one in the urban area could partly or fully use baby formula as a supplement for breastfeeding.

Geographical factors are observed to clearly play a role in child health. Compared to Nairobi province, we find the Coastal region, the Rift Valley and North Eastern provinces have lower *bmi's,* for both girls and boys. However, for the case of height for age and weight for age, Nyanza and North Eastern provinces are areas where a child is more likely to be taller and heavier compared to a child from Nairobi province. The likely explanation for this result is with the communities living in these areas. The North Eastern province is inhabited by Somali people and other Cushite people such as Rendille and the Afar. The area is mostly a remote semi-arid and arid area with extreme and severe climatic conditions with pronounced shortage of rains and therefore, life in the area is very harsh, lacking food and health facilities. Somalis seem to be quite tall relative to other Sub-Saharan Africans. These people are nomadic, pastoral people and have a culture primarily centred around camels with cattle and goats. Families live in portable huts which keep on moving in search of pasture and water for their cattle. Their diet consists almost entirely of milk and milk products and sometimes maize meal and rice. (see Guerin *et al.*, 2007). This feature could be the reason why these children are likely to have lower *bmi*.

Model 4 reveals that the female child in Nyanza province has higher values of both height for age and weight for age than those in Nairobi province. Again the possible reason for this could be the community itself. Nyanza province is occupied by Luo tribe from Nilotic group and is the third largest ethnic group with an estimated 13% of the population in Kenya. One dominant practice for the Luo is fishing given their close proximity to Lake Victoria. This makes fish their major staple food and "*ugali*", which is a very thick porridge, like cake, made with maize flour. Another popular dish is "*nyoyo",* which is a mixture of boiled maize and beans. This fish is used for both

domestic consumption and important economic activity where they export, especially the Nile Perch, to Europe and other countries. Comparatively, Luo people have big bodies and are usually tall (Roberts & Bainbridge, 1963).

The results suggest that wealth has no effect on *bmi* but does have very large positive effect on height for age and weight for age. Using *bmi* as summary measure of health, there is no clear wealth gradient for child health. This is an encouraging sign that even the poorest are able to maintain a reasonable level of health so that *bmi* does not vary with wealth. So how do we explain the strong positive effect of wealth on height for age and weight for age? Height partly reflects inherited stock of health from previous generations. Generations of poor nutrition lead to poorer health and shorter adults who in turn breed shorter children. So height for age captures long run (across generations) economic deprivation, at least in part. The fact that in the sample across wealth categories, weight is adequate for a given height for age to give *bmi*'s that do not vary with wealth, suggests that current levels of health and nutrition are adequate to ensure poorer children are not malnourished and underweight relative to their height for age. If this interpretation is correct, it is an encouraging picture of the health effects of wealth, at least in the short term.

It is very difficult for a poor mother to afford the cost of medical care, sometimes resulting in debts which may further deteriorate the family's living conditions. In fact, with persistent debts, it may take the family into a vicious circle of poverty. Such families cannot afford enough food and right diet for the child and other family members leading to child malnutrition. Earlier studies indicate poverty results in poor health which can contribute to child mortality. For example, according to Akhtar *et al.* (2005) in developing countries, people still seek health care from the traditional health care

practitioners which is a significant reason for poor maternal and child health and high incidence of child mortality and morbidity. In addition, based on the World Bank [WB] report (2006d), malnutrition is an important contributor to child mortality. However, our finding in this study confirms that in Kenya in 2003 wealth was not a contributing factor to child health, using *bmi* as the best measure of health.

Turning to the mother's characteristics, we took mother's education as one of the inputs or investments of the mother through which she is able to positively influence her child's health. The results are revealed by the dummy variables for different levels of education. Compared to the mother who did not have any primary school education, the results showed that mothers with at least some primary school education greatly improved their child's health in terms of *bmi* for age and weight for age. The possible reason why a mothers education boosts her child's health is because educated mothers are likely to be working, doing business and adopting the recommended modern medical and clinical facilities and lifestyle. They are able to know the correct nutritional diet for their children and able to immunise their children at the right time with a full dose. This is less likely to be the case for mothers who are not educated.

Our results match the previous studies carried out on similar grounds. For example, using Nigerian data, Caldwell (1979; 1994) was able to link the maternal education levels of mothers as a factor contributing to child mortality in Nigeria. The possible reason he gave to explain this effect was that mother's education works through changing feeding and care practices, leading to better health seeking behaviour and by changing the traditional family structure. Similarly, Hobcraft (1993), using DHS data from different countries (sub-Saharan African), found that maternal education was more likely to contribute to a child's survival by enabling women to have fewer children and

utilising pre-natal care and immunising their children. However, in his earlier studies, Hobcraft *et al.* (1984) had found that the effect in sub–Saharan Africa was weak. Also, in their study using 1991/92 Tanzania Demographic Health Surveys (TDHS), Mturi and Curtis (1995) found an insignificant effect of socioeconomic factors such as maternal education towards child mortality in the country.

Desai and Alva (1998) using data from Demographic and Health Surveys for 22 developing countries, examined the effect of maternal education on three markers of child health: infant mortality, children's height for age, and immunization status and they found a strong correlation between maternal education and these markers of child health in only a handful of countries, with education acting as a proxy for the socioeconomic status of the family and geographic area of residence. Chen and Li (2009), using a Chinese Children's Survey that was conducted by the National Bureau of Statistics of China in June 1992, examined the effect of maternal education on the health of young children and found that despite adopted children being genetically unrelated to the nurturing parents, the educational effect on these children was most likely to be the nurturing effect. They found that the mother's education is an important determinant of the health of adopted children even after they controlled for income, the number of siblings, health environments, and other socioeconomic variables.

A study by Kabubo-Mariara, *et al.* (2009) using a pooled sample of the 1998 and 2003 Demographic and Health Surveys data sets for Kenya found that male children suffer more malnutrition than females. The results further indicate that maternal education is a more important determinant of children's nutritional status than paternal education. Household assets are also important determinants of children's nutritional status but nutrition improves at a decreasing rate with assets.

The mother's health measured by *bmi,* also had a very strong effect on the child's health for all three indicators. Healthier mothers are more likely to give birth to a healthy child and the child is more likely to grow up healthy. The same results are found in the study done by Rahman *et al.* (1993). They found that the child's nutritional status, indicated by weight for age, was associated with the body mass index of the mother, and breastfeeding status of the child. This meant that maternal nutritional status is a proximate determinant of a child's nutritional status and they advise that it should be considered in programmes aiming at improving child health.

The mother's experience of violence with her husband or partner is another factor with a strong effect on child health. This is evident in the case of child's weight where the child whose mother is engaged in a violent relationship with her husband or partner, is likely to have lower weight, especially male children. There is some evidence of the same effect on the child's height. Violence in this case is considered as physical, verbal, sexual or emotional abuse. The experience of family violence can be among the most disturbing for children because both victims and aggressors are the adults who care for them and who are most closely attached to them. In any violence in the family, especially in an African setting, women tend to run away from the house for their safety leaving the young ones by themselves. For many of these children, violence interrupts their experience of consistent safety and care, and creates an environment of uncertainty and helplessness. These include problems with sleeping, eating and other basic bodily functions; depression, aggressiveness, anxiety and this affects the future of the child in regulating emotions; difficulties with family and peer relationships; and problems with attention, concentration and school performance (Hart & Brassard 1987). Based on Jaffe *et al*. (1990), infants are the most limited of all children in their cognitive abilities and

resources for adaptation. Infants who witness spousal violence are often characterised by poor health, poor sleeping habits and excessive screaming.

As for other maternal characteristics there is only weak evidence of household size or number of births in the last 5 years having a detrimental effect on *bmi* for age, height for age or weight for age. A stronger effect might have been expected, given common sense and prior research. With restricted household resources or wealth, more children reduces the chances of each child receiving suitable medical care and even a proper food diet. Every time a mother gets pregnant her productivity in terms of labour and therefore income is affected. As a result, giving birth to more children within five years affects the child even more severely in terms of its health (WHO, 2009). Most previous studies on child health, or mortality rates in developing countries with consideration of birth spacing show that child spacing or birth interval is one of the factors influencing infant health as well as other childhood stages of life. Mothers with short intervals between two pregnancies have insufficient time to restore their nutritional reserves and as a result, this affects foetal growth. In their study, Mondal *et al.* (2009) found that the length of the preceding birth interval had very strong impact on infant and child survival.

Lastly, we need to mention that the effect of the mother being HIV positive does not come up significant for any of the child health indicators. However, it plays a big role in a child's survival in the selection equation. This is expected since the mother who is infected with the disease is likely to pass the disease on to the child and hence the chances for this child to survive are reduced. This could be exacerbated by the fact that in developing countries, medical facilities and access to treatment is often a problem.

The results on child survival give evidence in relation to a mothers' investment in their child's survival. This is best seen in Model 3. Relevant factors to survival are the mother's education, contraceptive use, tetanus injection given to the mother during pregnancy and pre-natal clinic visits. By way of comparison, Akhtar *et al*. (2005) study factors affecting child health with reference to rural Faisalabad and found that the delivery of a baby, mostly at home, where there are no trained birth attendants, was to a great extent a responsible factor for the high incidence of morbidity among children. They gave several suggestions after their findings and one of the suggestions was the desperate need to make women aware of the importance of pre-delivery and post-delivery medical checkup along with the delivery of babies under the supervision of trained medical health care practitioners (Akhtar *et al.;* 2005).

In this study, interestingly, we found higher mother's *bmi* has a negative effect on child survival. This is surprising, since BMI can be used as a good measure of nutritional status and health of adults, and healthier adults are more likely to have healthier children. One possible explanation is found in the Perinatal Mortality 2009 report (2011), where they find that the number of deaths in newborn babies is high where the mother's *bmi* value is very high. The report found that although there has been a downward trend of neonatal mortality rates over the last decade, 10% of mothers who had a stillbirth had a BMI of 35 or more.

As mentioned above, the variable for whether the mother is infected with HIV was included in the survival equation. We found HIV to be one of the contributing factors to the survival of the child, especially the male child. This means that a child born to a mother with HIV disease is more likely to die. Child survival is influenced by the HIV epidemic through several mechanisms. Firstly, the disease is often passed on to the child

and eventually kills the child. In most cases, the mother-to-child transmission for HIV infection is rare at early pregnancy but relatively frequent in late pregnancy and during delivery. Breastfeeding also contributes substantially to the overall risk. For most babies, breastfeeding is seen as the best way to be fed. If no antiretroviral drugs are being taken, however, breastfeeding for two or more years can double the risk of the baby becoming infected to around 40% (de Cock, *et al.*, 2000; Fowler, *et al*., 2002). In sub-Saharan Africa, HIV infection is an ever-increasing cause of child mortality, although the effect varies from country to country. There is a greater effect in southern Africa where HIV causes up to half of all child deaths in the worst affected countries (Newell, Brahmbhatt and Ghys, 2004). In their attempt to study the causes for a 25% increase in under-five mortality in Kenya between the late 1980s and mid 1990s, Kenneth and his fellow researchers found that HIV prevalence in a community was found to be associated with an increase in child mortality. They included a number of control variables in the analyses including social, bio-demographic and health sector factors and they found that even while controlling for these variables, HIV was found to be significant suggesting that its direct effect was substantial (Hill, *et al*., 2001).

This study has sought to give insight into child health assessment, focusing on factors mostly responsible for health status. From this study, child health policymakers can observe the factors which seem to affect child health and can design strategies which improve the health of children in Kenya. The study also shows some factors which can help the child survive, particularly with improvement in the mother's investment in their own and their child's health and access to health facilities, particularly around delivery of the child.

Table 4.2. Descriptive Statistic for Selected Variables

Dependent Variable	All		Urban		Rural		Female		Male	
	Sample size	Proportion/ Mean(Std dev)	Sample Size (%)	Proportion/ Mean(Std dev)	Sample Size (%)	Proportion/ Mean(Std dev)	Sample Size (%)	Proportion/ Mean(Std dev)	Sample Size (%)	Proportion/ Mean(Std dev)
Child Characteristics: Physical										
Alive child	5949	0.916	25.8%	0.919	74.21%	0.914	49.32%	0.927	50.68%	0.905
		(0.278)		(0.273)		(0.280)		(0.261)		(0.294)
Child weight	4957	11.011	24.3%	11.311	75.70%	10.939	50.06%	10.869	49.94%	11.190
		(3.561)		(3.769)		(3.771)		(4.026)		(3.495)
Child height	4887	82.373	24.1%	82.945	75.85%	82.191	49.93%	82.070	50.07%	82.675
		(14.649)		(14.975)		(14.542)		(15.003)		(14.285)
Body mass index (bmi)	4887	16.162	24.1%	16.443	75.85%	16.072	49.93%	16.088	50.07%	16.236
		(3.693)		(3.904)		(3.619)		(4.483)		(2.682)
Child Size at birth:										
born very large *(base)*	5931	0.052	25.8%	0.069	74.19%	0.046	49.33%	0.049	50.67%	0.055
		(0.222)		(0.254)		(0.209)		(0.216)		(0.227)
born larger than average	5931	0.191	25.8%	0.169	74.19%	0.199	49.33%	0.168	50.67%	0.214
		(0.393)		(0.375)		(0.399)		(0.374)		(0.410)
born average	5931	0.587	25.8%	0.605	74.19%	0.581	49.33%	0.590	50.67%	0.584
		(0.492)		(0.489)		(0.493)		(0.492)		(0.493)
born smaller than average	5931	0.123	25.8%	0.093	74.19%	0.133	49.33%	0.135	50.67%	0.111
		(0.328)		(0.291)		(0.340)		(0.342)		(0.314)
born very small	5931	0.043	25.8%	0.057	74.19%	0.038	49.33%	0.054	50.67%	0.033
		(0.203)		(0.232)		(0.192)		(0.225)		(0.178)
Child born twins	5949	0.034	25.8%	0.035	74.21%	0.034	49.32%	0.038	50.68%	0.030
		(0.181)		(0.183)		(0.180)		(0.192)		(0.169)
Child Characteristics: Health										
Months breastfed	5679	2.761	25.6%	2.610	74.45%	2.813	49.52%	2.733	50.48%	2.789
		(1.245)		(1.241)		(1.242)		(1.230)		(1.259)
Given water	4687	0.793	26.1%	0.843	73.94%	0.775	50.23%	0.794	49.77%	0.792
		(0.405)		(0.364)		(0.417)		(0.404)		(0.406)
Given juice	4687	0.191	26.0%	0.324	73.98%	0.145	50.31%	0.187	49.69%	0.195
		(0.393)		(0.468)		(0.352)		(0.390)		(0.397)

Table 4.2: ***Continued***

Dependent Variable	All		Urban		Rural		Female		Male	
	Sample size	Proportions/ Mean (Std dev)	Sample Size (%)	Proportions/ Mean (Std dev)	Sample Size (%)	Proportions/ Mean (Std dev)	Sample Size (%)	Proportions/ Mean (Std dev)	Sample Size (%)	Proportions/ Mean (Std dev)
Given commercially produced baby formula	4687	0.048 (0.213)	26.0%	0.096 (0.295)	73.99%	0.031 (0.173)	50.29%	0.048 (0.213)	49.71%	0.048 (0.214)
Given powdered/fresh animal milk	4687	0.554 (0.497)	26.0%	0.558 (0.497)	74.03%	0.553 (0.497)	50.31%	0.557 (0.497)	49.69%	0.552 (0.497)
Given pumpkin, carrots, red/yel yams, red sweet pot	4687	0.185 (0.388)	26.0%	0.260 (0.439)	74.01%	0.158 (0.365)	50.31%	0.178 (0.383)	49.69%	0.191 (0.393)
Given green vegetables	4687	0.494 (0.500)	26.0%	0.484 (0.500)	74.01%	0.498 (0.500)	50.31%	0.494 (0.500)	49.69%	0.494 (0.500)
Given vitamin A rich fruits, eg mango, papaya	4687	0.218 (0.413)	26.0%	0.292 (0.455)	74.01%	0.192 (0.394)	50.31%	0.214 (0.410)	49.69%	0.222 (0.416)
Given food made from local grain	4686	0.660 (0.474)	26.0%	0.670 (0.470)	74.04%	0.656 (0.475)	50.29%	0.656 (0.475)	49.71%	0.664 (0.473)
Child Characteristics: Vaccination										
Child received bcg	5445	0.841 (0.366)	25.9%	0.897 (0.304)	74.10%	0.821 (0.383)	49.90%	0.840 (0.367)	50.10%	0.842 (0.365)
Child received dpt	5445	0.835 (0.371)	25.9%	0.867 (0.340)	74.10%	0.825 (0.380)	49.90%	0.827 (0.378)	50.10%	0.843 (0.363)
Child received polio	5442	0.851 (0.356)	25.9%	0.877 (0.329)	74.09%	0.842 (0.365)	49.91%	0.844 (0.363)	50.09%	0.858 (0.349)
Child's mothers Characteristics: Wealth (index)										
From poorest household *(base)*	5949	0.252 (0.434)	25.8%	0.048 (0.214)	74.21%	0.323 (0.468)	49.32%	0.253 (0.435)	50.68%	0.251 (0.434)
From poorer household	5949	0.188 (0.391)	25.8%	0.038 (0.192)	74.21%	0.240 (0.427)	49.32%	0.197 (0.398)	50.68%	0.179 (0.384)
From middle household	5949	0.181 (0.385)	25.8%	0.061 (0.240)	74.21%	0.223 (0.416)	49.32%	0.178 (0.382)	50.68%	0.184 (0.388)
From Richer household	5949	0.158 (0.364)	25.8%	0.127 (0.333)	74.21%	0.168 (0.374)	49.32%	0.152 (0.359)	50.68%	0.163 (0.370)
From Richest household	5949	0.222 (0.415)	25.8%	0.725 (0.447)	74.21%	0.047 (0.211)	49.32%	0.221 (0.415)	50.68%	0.222 (0.416)
Wealth, Continuous	5896	-0.116 (3.031)	25.9%	0.278 (3.370)	74.10%	-0.254 (2.891)	49.36%	-0.162 (2.912)	50.64%	-0.072 (3.142)
Wealth Continuous residuals	5870	-0.012 (2.983)	25.8%	-0.042 (3.391)	74.22%	-0.001 (2.829)	49.44%	-0.056 (2.870)	50.56%	0.031 (3.090)

Table 4.2: ***Continued***

Dependent Variable	All		Urban		Rural		Female		Male	
	Sample size	**Proportions/ Mean (Std dev)**	**Sample Size (%)**	**Proportions/ Mean (Std dev)**	**Sample Size (%)**	**Proportions/ Mean (Std dev)**	**Sample Size (%)**	**Proportions/ Mean (Std dev)**	**Sample Size (%)**	**Proportions/ Mean (Std dev)**
Child's mothers Characteristics: ***Residence/Region***										
Urban	5949	0.258 (0.437)	25.8%				49.32%	0.256 (0.436)	50.68%	0.260 (0.439)
Province:										
Nairobi (Base)	5949	0.092 (0.289)	25.8%	0.357 (0.479)	74.21%	0.000 (0.000)	49.32%	0.093 (0.290)	50.68%	0.092 (0.288)
Central	5949	0.123 (0.328)	25.8%	0.074 (0.261)	74.21%	0.140 (0.347)	49.32%	0.122 (0.327)	50.68%	0.124 (0.329)
Coast	5949	0.117 (0.322)	25.8%	0.154 (0.362)	74.21%	0.105 (0.306)	49.32%	0.112 (0.316)	50.68%	0.123 (0.328)
Eastern	5949	0.118 (0.322)	25.8%	0.023 (0.149)	74.21%	0.151 (0.358)	49.32%	0.113 (0.316)	50.68%	0.122 (0.328)
Nyanza	5949	0.133 (0.340)	25.8%	0.093 (0.290)	74.21%	0.147 (0.354)	49.32%	0.137 (0.344)	50.68%	0.130 (0.336)
Rift Valley	5949	0.202 (0.401)	25.8%	0.110 (0.313)	74.21%	0.234 (0.423)	49.32%	0.205 (0.404)	50.68%	0.198 (0.399)
Western	5949	0.139 (0.346)	25.8%	0.114 (0.318)	74.21%	0.148 (0.355)	49.32%	0.145 (0.352)	50.68%	0.134 (0.341)
North Eastern	5949	0.076 (0.265)	25.8%	0.075 (0.263)	74.21%	0.076 (0.266)	49.32%	0.074 (0.262)	50.68%	0.078 (0.268)
Child's mothers Characteristics: Education status										
No education	5949	0.203 (0.403)	25.8%	0.143 (0.350)	74.21%	0.224 (0.417)	49.32%	0.200 (0.400)	50.68%	0.207 (0.405)
With at least primary education	5949	0.581 (0.493)	25.8%	0.490 (0.500)	74.21%	0.613 (0.487)	49.32%	0.588 (0.492)	50.68%	0.574 (0.495)
With at least secondary or higher education	5949	0.216 (0.411)	25.8%	0.368 (0.482)	74.21%	0.163 (0.369)	49.32%	0.212 (0.409)	50.68%	0.219 (0.414)
Child's mothers Characteristics: ***Others Variables***										
Violence	5936	0.346 (0.476)	25.7%	0.308 (0.462)	74.28%	0.359 (0.480)	49.36%	0.345 (0.476)	50.64%	0.347 (0.476)
Mother HIV/AIDS positive	2448	0.082 (0.274)	22.9%	0.123 (0.329)	77.08%	0.069 (0.254)	49.47%	0.093 (0.291)	50.53%	0.070 (0.256)
Imputed HIV/AIDS mothers	4749	0.086 (0.281)	24.6%	0.128 (0.334)	75.41%	0.086 (0.280)	49.21%	0.085 (0.279)	50.79%	0.078 (0.269)

Table 4.3. OLS and Heckman Estimates for child's Body Mass Index (bmi) for age (in months)

Dependent Variable: z – score for *bmi* for age[53]					
	OLS	***Heckman's two-step consistent estimates***			
Independent Variables	**Model 1** No. obs 1,917	**Model 2** No. obs 3,971 Censd obs 363 Uncensd obs 3,608	**Model 3** No. obs 3,845 Censd obs 343 Uncensd obs 3,502	**Model 4** No. obs 1,897 Censored obs 141 Uncensored obs 1,756	**Model 5** No. obs 1,948 Censored obs 202 Uncensored obs 1,746
Child Characteristics					
Child sex (*Base - female*)	-0.011	-0.063	-0.057		
	(0.061)	(0.045)	(0.046)		
Child age - all	-0.278***	-0.273***	-0.281***	-0.253***	-0.313***
	(0.037)	(0.028)	(0.028)	(0.040)	(0.041)
Child age - above 6	0.308***	0.303***	0.313***	0.281***	0.348***
	(0.043)	(0.033)	(0.033)	(0.047)	(0.048)
Child age - above 18	-0.033**	-0.031***	-0.033***	-0.030*	-0.039**
	(0.015)	(0.011)	(0.011)	(0.016)	(0.016)
Child age - above 54 up to 59	-0.044	-0.032	-0.026	-0.061	0.016
	(0.044)	(0.031)	(0.032)	(0.045)	(0.044)
Child born larger than average (*Base – largest*)	-0.089	-0.017	-0.028	-0.213	0.185
	(0.157)	(0.110)	(0.111)	(0.162)	(0.152)
Child born average	-0.370**	-0.303***	-0.310***	-0.582***	-0.024
	(0.146)	(0.102)	(0.103)	(0.149)	(0.142)
Child born smaller than average	-0.587***	-0.565***	-0.547***	-0.757***	-0.326*
	(0.166)	(0.119)	(0.120)	(0.170)	(0.168)
Child born very small	-0.579***	-0.442***	-0.431***	-0.639***	-0.221
	(0.202)	(0.150)	(0.151)	(0.202)	(0.234)
Child born twins	-0.238	-0.174	-0.120	-0.320*	0.076
	(0.187)	(0.144)	(0.145)	(0.194)	(0.218)
Child received bcg vaccine	-0.184	-0.125	-0.124	-0.227*	-0.057
	(0.136)	(0.098)	(0.099)	(0.132)	(0.148)
Child received dpt vaccine	0.035	0.089	0.074	-0.027	0.230
	(0.173)	(0.119)	(0.120)	(0.160)	(0.183)

[53] ***Models 1***: Regression estimates for z-score for *bmi*-for-age using ordinal household wealth variable (no endogeneity bias correction); HIV/AIDS variable (only observed data – sample selection ignored) and child being born alive selection ignored. ***Models 2:*** Heckman estimates for z-score for *bmi*-for-age using ordinal wealth variable (no endogeneity bias correction); imputed HIV/AIDS variable and selectivity Model for child being born alive or dead. ***Model 3:*** same as **Model 2**, but wealth endogeneity corrected by construction of wealth continuous variable and residuals manually constructed. ***Model 4:*** same as **Model 4 and 5** but for only female and male child separately.

Table 4.3: ***Continued***

Dependent Variable: z – score for *bmi* for age

	OLS		***Heckman's two-step consistent estimates***		
Independent Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Child received polio vaccine	0.196	-0.011	-0.000	0.141	-0.149
	(0.159)	(0.112)	(0.114)	(0.158)	(0.165)
Child given vit. A last 6 months	0.005	0.028	0.027	0.000	0.024
	(0.068)	(0.050)	(0.051)	(0.073)	(0.071)
Months breastfed	0.010	0.007	0.018	0.020	0.023
	(0.036)	(0.027)	(0.027)	(0.039)	(0.038)
Child given water	-0.073	-0.008	-0.036	-0.013	-0.072
	(0.099)	(0.072)	(0.073)	(0.103)	(0.104)
Child given juice	-0.075	-0.121	-0.092	-0.056	-0.134
	(0.119)	(0.085)	(0.087)	(0.125)	(0.120)
Child given commercially produced baby formula	-0.363*	-0.062	-0.069	-0.157	-0.001
	(0.216)	(0.164)	(0.168)	(0.256)	(0.223)
Child given powdered/fresh animal milk	-0.179**	-0.097	-0.095	-0.065	-0.149*
	(0.086)	(0.063)	(0.064)	(0.090)	(0.089)
Child given pumpkin, carrots, red/yel yams, red sweet potatoes	-0.056	-0.005	-0.015	0.026	-0.049
	(0.117)	(0.084)	(0.085)	(0.123)	(0.119)
Child given green vegetables	0.230**	0.084	0.089	-0.095	0.269***
	(0.095)	(0.069)	(0.071)	(0.101)	(0.099)
Child given Vit. A rich fruits, eg. mango, papaya	0.134	0.072	0.067	0.178	-0.020
	(0.109)	(0.079)	(0.080)	(0.115)	(0.112)
Child given food made from local grain	-0.073	-0.001	0.010	0.113	-0.103
	(0.095)	(0.071)	(0.072)	(0.102)	(0.102)
Household Characteristics					
Number of household members	-0.012	-0.001	0.001	0.001	-0.004
	(0.014)	(0.013)	(0.013)	(0.017)	(0.018)
Residence/Region					
Urban *(Base - rural)*	-0.031	0.029	0.047	-0.020	0.115
	(0.108)	(0.084)	(0.089)	(0.128)	(0.123)
Central province *(Base - Nairobi)*	0.006	0.030	0.023	-0.054	0.081
	(0.171)	(0.124)	(0.122)	(0.172)	(0.173)
Coast province	-0.314*	-0.261**	-0.274**	-0.218	-0.306
	(0.171)	(0.120)	(0.136)	(0.193)	(0.191)
Eastern province	-0.359**	-0.169	-0.167	-0.235	-0.105
	(0.181)	(0.128)	(0.132)	(0.188)	(0.187)
Nyanza province	-0.174	-0.045	-0.073	-0.085	-0.069
	(0.171)	(0.123)	(0.131)	(0.188)	(0.184)

Table 4.3: ***Continued***

Dependent Variable: z – score for *bmi* for age

	OLS		***Heckman's two-step consistent estimates***		
Independent Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Rift Valley province	-0.480***	-0.323***	-0.321***	-0.286*	-0.331**
	(0.167)	(0.116)	(0.115)	(0.164)	(0.162)
Western province	-0.290*	-0.122	-0.147	-0.215	-0.100
	(0.174)	(0.121)	(0.129)	(0.180)	(0.184)
North Eastern province	-0.829***	-0.825***	-0.820***	-1.024***	-0.544**
	(0.226)	(0.163)	(0.191)	(0.267)	(0.272)
Wealth level					
Poorest (*Base – richest*)	0.132	-0.000			
	(0.144)	(0.108)			
Poorer	0.026	-0.097			
	(0.140)	(0.105)			
Medium	0.099	-0.031			
	(0.137)	(0.102)			
Richer	0.061	-0.072			
	(0.131)	(0.097)			
Wealth, Continuous			0.000	0.034	-0.038
			(0.199)	(0.287)	(0.274)
Wealth Continuous residuals			-0.013	-0.085	0.062
			(0.207)	(0.303)	(0.282)
Mothers Characteristics					
Mother's Education level					
With at least primary education (*Base - no education*)	0.265***	0.302***	0.296***	0.249**	0.339***
	(0.101)	(0.079)	(0.080)	(0.114)	(0.112)
With at least secondary or higher education	0.431***	0.536***	0.555***	0.415***	0.633***
	(0.123)	(0.099)	(0.102)	(0.145)	(0.143)
Others					
HIV/AIDS positive	0.079				
	(0.119)				
Imputed HIV/AIDS mothers		0.004	-0.007	-0.071	0.096
		(0.088)	(0.091)	(0.122)	(0.132)
Births in last 5yrs	0.014	-0.045	-0.040	0.007	-0.067
	(0.050)	(0.038)	(0.038)	(0.053)	(0.054)
Mother's *bmi*	0.001***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 4.3: ***Continued***

Dependent Variable: z – score for *bmi* for age					
	OLS	***Heckman's two-step consistent estimates***			
Independent Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Violence with partner	0.003	-0.063	-0.071	-0.019	-0.119
	(0.067)	(0.050)	(0.053)	(0.074)	(0.074)
Inverse Mills Ratio		0.512*	0.670**	0.865**	0.305
		(0.299)	(0.307)	(0.391)	(0.376)
Constant	0.201	0.615**	0.592**	0.418	0.621
	(0.354)	(0.262)	(0.278)	(0.415)	(0.389)

Standard errors are in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%

Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 4.4. Heckman Selection Model for child being alive or dead (for bmi)

	Heckman's two-step consistent estimates			
	Model 2	**Model 3**	**Model 4**	**Model 5**
Household Characteristics				
Number of Household members[54]	0.122***	0.126***	0.106***	0.154***
	(0.016)	(0.016)	(0.025)	(0.022)
Wealth level				
Poorest	0.036			
	(0.136)			
Poorer	0.037			
	(0.135)			
Middle	-0.037			
	(0.133)			
Richer	-0.050			
	(0.124)			
Wealth Continuous		0.045	0.651	-0.301
		(0.278)	(0.427)	(0.387)
Wealth Continuous residuals		-0.142	-0.783*	0.261
		(0.286)	(0.443)	(0.396)
Mother's Characteristics				
Mother's education level				
With at least primary education	0.064	0.120	0.352**	-0.062
	(0.111)	(0.113)	(0.168)	(0.158)
With at least secondary or higher education	0.276**	0.273*	0.370	0.283
	(0.140)	(0.161)	(0.236)	(0.233)
Mother's Other Variables				
Age when had first sex intercourse	-0.124*	-0.124*	0.003	-0.194**
	(0.071)	(0.073)	(0.115)	(0.098)
Use physical contraception	0.206	0.175	0.165	0.204
	(0.189)	(0.190)	(0.281)	(0.273)
Use hormonal contraception	0.157**	0.159**	0.089	0.274**
	(0.078)	(0.080)	(0.119)	(0.113)

[54] See Table 6 for illustration of the effect of household size on a child's survival

Table 4.4: ***Continued***

	Heckman's two-step consistent estimates			
	Model 2	**Model 3**	**Model 4**	**Model 5**
Mother's *bmi*	-0.000***	-0.000***	-0.001***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Mother given tetanus injection	0.248*	0.246*	0.610**	-0.052
	(0.144)	(0.149)	(0.238)	(0.203)
Pre-natal clinic visits (doctor/nurse/midwife)	0.333**	0.360**	0.176	0.554***
	(0.147)	(0.153)	(0.245)	(0.207)
Amenorrhea period	0.023	0.027	0.013	0.040
	(0.030)	(0.031)	(0.048)	(0.043)
Mother took iron tabs when pregnant	0.025	0.029	0.026	0.059
	(0.088)	(0.092)	(0.150)	(0.121)
Delivered at hospital	-0.095	-0.033	-0.038	-0.001
	(0.074)	(0.078)	(0.122)	(0.105)
Delivered by caesarean	-0.130	-0.189	-0.398	0.001
	(0.168)	(0.170)	(0.252)	(0.239)
Urban	0.048	0.005	0.047	-0.025
	(0.107)	(0.122)	(0.189)	(0.169)
Imputed HIV/AIDS mothers	-0.156	-0.185*	-0.140	-0.268*
	(0.102)	(0.104)	(0.159)	(0.144)
Violence with partner	-0.123*	-0.146**	-0.119	-0.201**
	(0.063)	(0.068)	(0.103)	(0.094)
Mother's tribe				
Kikuyu (*Base Embu & others*)	0.320*	0.260	0.418*	0.102
	(0.166)	(0.168)	(0.247)	(0.241)
Kalenjin	-0.053	-0.073	0.052	-0.224
	(0.167)	(0.170)	(0.244)	(0.248)
Kamba	-0.105	-0.110	-0.043	-0.227
	(0.170)	(0.173)	(0.247)	(0.253)
Kisii	-0.134	-0.087	-0.184	-0.029
	(0.190)	(0.200)	(0.273)	(0.305)
Luhya	-0.095	-0.076	0.089	-0.277
	(0.153)	(0.156)	(0.224)	(0.228)
Luo	-0.420***	-0.406**	-0.540**	-0.378
	(0.158)	(0.160)	(0.229)	(0.233)
Masai	0.342	0.341	0.282	0.241
	(0.241)	(0.277)	(0.446)	(0.374)
Meru	0.284	0.214	0.140	0.163
	(0.250)	(0.251)	(0.352)	(0.369)

Table 4.4: ***Continued***

	Heckman's two-step consistent estimates			
	Model 2	**Model 3**	**Model 4**	**Model 5**
Mijikenda/Swahili	-0.138	-0.119	0.497*	-0.533**
	(0.172)	(0.181)	(0.297)	(0.257)
Somali	-0.328*	-0.234	0.350	-0.703***
	(0.170)	(0.178)	(0.263)	(0.257)
Constant	0.943***	0.863***	2.007***	0.398
	(0.243)	(0.313)	(0.521)	(0.433)
Heckman maximum likelihood estimates				
ρ	0.328	0.370	0.500	0.344
	(0.115)	(0.104)	(0.103)	(0.194)
σ	1.373	1.372	1.378	1.358
	(0.021)	(0.021)	(0.028)	(0.033)
λ (inverse Mills ratio)	0.450	0.508	0.690	0.468
	(0.163)	(0.147)	(0.150)	(0.271)
LR test of independent equations ($\rho = 0$)	$\chi_1^2 = 3.900$	$\chi_1^2 = 5.680$	$\chi_1^2 = 7.88$	$\chi_1^2 = 1.450$
	$Prob > \chi^2 = 0.048$	$Prob > \chi^2 = 0.017$	$Prob > \chi^2 = 0.005$	$Prob > \chi^2 = 0.228$

Standard errors are in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%

Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 4.5. OLS and Heckman Estimates for child's height for age (in months)

Dependent Variable: z – score for height-for-age					
	OLS	***Heckman's two-step consistent estimates***			
Independent Variables	**Model 1** No. obs 1,908	**Model 2** No. obs 3,961 Censored obs 363 Uncensored obs 3,598	**Model 3** No. obs 3,836 Censored obs 343 Uncensored obs 3,493	**Model 4** No. obs 1,892 Censored obs 141 Uncensored obs 1,781	**Model 5** No. obs 1,944 Censored obs 204 Uncensored obs 1,742
Child Characteristics					
Child sex (*Base - female*)	-0.258***	-0.174***	-0.169***		
	(0.068)	(0.048)	(0.049)		
Child age - all	-0.050	-0.047	-0.056*	-0.030	-0.102**
	(0.041)	(0.030)	(0.030)	(0.042)	(0.043)
Child age - above 6	-0.045	-0.064*	-0.055	-0.073	-0.015
	(0.048)	(0.035)	(0.036)	(0.050)	(0.050)
Child age - above 18	0.112***	0.132***	0.133***	0.123***	0.142***
	(0.016)	(0.012)	(0.012)	(0.017)	(0.017)
Child age - above 54 up to 59	-0.031	-0.053	-0.054	-0.014	-0.090*
	(0.048)	(0.033)	(0.033)	(0.048)	(0.047)
Child born larger than average (*Base – largest*)	-0.014	-0.048	-0.034	-0.068	-0.007
	(0.174)	(0.117)	(0.118)	(0.172)	(0.161)
Child born average	-0.301*	-0.281***	-0.282**	-0.229	-0.332**
	(0.162)	(0.109)	(0.110)	(0.159)	(0.150)
Child born smaller than average	-0.712***	-0.493***	-0.492***	-0.523***	-0.393**
	(0.184)	(0.126)	(0.128)	(0.181)	(0.179)
Child born very small	-0.629***	-0.694***	-0.752***	-0.568***	-1.153***
	(0.224)	(0.160)	(0.162)	(0.216)	(0.251)
Child born twins	-0.300	-0.747***	-0.698***	-0.483**	-0.978***
	(0.205)	(0.153)	(0.155)	(0.207)	(0.232)
Child received bcg vaccine	0.238	-0.050	-0.034	-0.084	-0.006
	(0.152)	(0.103)	(0.105)	(0.139)	(0.158)
Child received dpt vaccine	0.331*	0.221*	0.252**	0.323*	0.189
	(0.191)	(0.125)	(0.128)	(0.170)	(0.194)
Child received polio vaccine	-0.174	0.138	0.119	-0.058	0.343**
	(0.175)	(0.117)	(0.119)	(0.165)	(0.174)
Child given vit. A last 6 months	0.101	0.074	0.078	0.095	0.075
	(0.075)	(0.053)	(0.054)	(0.077)	(0.076)

Table 4.5: ***Continued***

	OLS		*Heckman's two-step consistent estimates*		
Independent Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Child Characteristics					
Months breastfed	-0.120***	-0.130***	-0.131***	-0.131***	-0.141***
	(0.040)	(0.029)	(0.029)	(0.042)	(0.041)
Child given water	0.131	-0.073	-0.088	0.003	-0.151
	(0.109)	(0.077)	(0.078)	(0.108)	(0.110)
Child given juice	0.084	0.051	0.025	-0.056	0.134
	(0.132)	(0.091)	(0.093)	(0.131)	(0.130)
Child given commercially produced baby formula	0.819***	0.450***	0.456**	0.613**	0.292
	(0.234)	(0.172)	(0.177)	(0.257)	(0.242)
Child given powdered/fresh animal milk	0.153	0.068	0.048	-0.114	0.232**
	(0.095)	(0.067)	(0.068)	(0.095)	(0.095)
Child given pumpkin, carrots, red/yel yams, red sweet potatoes	0.185	0.148	0.124	0.036	0.237*
	(0.129)	(0.090)	(0.092)	(0.129)	(0.128)
Child given green vegetables	-0.124	-0.083	-0.044	0.212**	-0.295***
	(0.105)	(0.074)	(0.076)	(0.106)	(0.106)
Child given Vit. A rich fruits, eg. mango, papaya	-0.101	0.066	0.061	-0.011	0.116
	(0.120)	(0.085)	(0.086)	(0.121)	(0.120)
Child given food made from local grain	-0.075	0.033	0.049	-0.091	0.211**
	(0.104)	(0.075)	(0.077)	(0.107)	(0.108)
Household Characteristics					
Number of household members	0.003	-0.017	-0.027*	-0.021	-0.020
	(0.015)	(0.014)	(0.014)	(0.018)	(0.020)
Residence/Region					
Urban *(Base - rural)*	0.075	-0.088	-0.054	-0.205	0.074
	(0.120)	(0.093)	(0.098)	(0.135)	(0.134)
Central province *(Base - Nairobi)*	0.170	-0.065	-0.011	0.025	0.042
	(0.189)	(0.132)	(0.131)	(0.180)	(0.185)
Coast province	0.258	0.063	0.207	0.134	0.313
	(0.189)	(0.128)	(0.146)	(0.202)	(0.205)
Eastern province	0.090	-0.087	0.009	0.048	0.057
	(0.201)	(0.137)	(0.142)	(0.197)	(0.200)
Nyanza province	0.503***	0.298**	0.351**	0.495**	0.229
	(0.189)	(0.131)	(0.141)	(0.198)	(0.197)
Rift Valley province	0.333*	0.058	0.022	0.119	-0.025
	(0.184)	(0.123)	(0.123)	(0.173)	(0.173)
Western province	0.363*	0.111	0.171	0.299	0.114
	(0.192)	(0.128)	(0.138)	(0.189)	(0.196)

Table 4.5: ***Continued***

	OLS		***Heckman's two-step consistent estimates***		
Independent Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
North Eastern province	0.907***	1.062***	1.289***	1.409***	1.203***
	(0.250)	(0.172)	(0.203)	(0.278)	(0.291)
Wealth level					
Poorest (*Base – richest*)	-0.540***	-0.611***			
	(0.159)	(0.119)			
Poorer	-0.297*	-0.338***			
	(0.154)	(0.116)			
Medium	-0.212	-0.245**			
	(0.151)	(0.112)			
Richer	-0.197	-0.232**			
	(0.145)	(0.108)			
Wealth, Continuous			0.771***	1.255***	0.383
			(0.217)	(0.302)	(0.297)
Wealth Continuous residuals			-0.455**	-0.834***	-0.236
			(0.226)	(0.318)	(0.305)
Mothers Characteristics					
Education level					
With at least primary education (*Base - no education*)	-0.224**	-0.238***	-0.110	-0.148	-0.093
	(0.113)	(0.087)	(0.088)	(0.121)	(0.121)
With at least secondary or higher education	0.109	0.045	0.089	0.036	0.115
	(0.136)	(0.109)	(0.112)	(0.153)	(0.155)
Others					
HIV/AIDS positive	0.075				
	(0.132)				
Imputed HIV/AIDS mothers		0.113	0.129	0.124	0.092
		(0.100)	(0.101)	(0.131)	(0.147)
Births in last 5yrs	-0.153***	-0.012	-0.030	-0.064	-0.008
	(0.056)	(0.040)	(0.041)	(0.056)	(0.058)
Mother *bmi*	0.000***	0.000***	0.000***	0.001***	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Violence with partner	-0.194***	-0.124**	-0.091	-0.082	-0.114
	(0.073)	(0.055)	(0.058)	(0.078)	(0.080)
Inverse Mills Ratio		-1.184***	-1.153***	-0.839**	-0.740*
		(0.321)	(0.328)	(0.414)	(0.397)
Constant	-0.327	0.170	0.346	-0.097	0.299
	(0.392)	(0.283)	(0.301)	(0.438)	(0.418)

Standard errors are in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%

Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 4.6. Heckman Selection Model for child being alive or dead (for Height)

	Heckman's two-step consistent estimates			
	Model 2	**Model 3**	**Model 4**	**Model 5**
Household Characteristics				
Number of Household members	0.122***	0.125***	0.104***	0.155***
	(0.016)	(0.016)	(0.025)	(0.022)
Wealth level				
Poorest	0.021			
	(0.136)			
Poorer	0.036			
	(0.135)			
Middle	-0.042			
	(0.133)			
Richer	-0.064			
	(0.124)			
Wealth Continuous		-0.007	0.643	-0.394
		(0.279)	(0.428)	(0.391)
Wealth Continuous residuals		-0.091	-0.772*	0.345
		(0.287)	(0.444)	(0.400)
Mother's Characteristics				
Mother's education level				
With at least primary education	0.067	0.124	0.358**	-0.057
	(0.111)	(0.113)	(0.169)	(0.159)
With at least secondary or higher education	0.269*	0.278*	0.381	0.284
	(0.140)	(0.162)	(0.237)	(0.233)
Others Variables				
Age when had first sex intercourse	-0.124*	-0.125*	0.009	-0.203**
	(0.071)	(0.073)	(0.115)	(0.098)
Use physical contraception	0.203	0.174	0.183	0.193
	(0.189)	(0.189)	(0.281)	(0.271)
Use hormonal contraception	0.157**	0.158**	0.092	0.268**
	(0.077)	(0.080)	(0.119)	(0.113)
Mother's *bmi*	-0.000***	-0.000***	-0.001***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Mother given tetanus injection	0.236*	0.233	0.595**	-0.058
	(0.143)	(0.149)	(0.236)	(0.203)

Table 4.6: ***Continued***

	Heckman's two-step consistent estimates			
	Model 2	**Model 3**	**Model 4**	**Model 5**
Pre-natal clinic visits (doctor/nurse/midwife)	0.347**	0.376**	0.199	0.559***
	(0.147)	(0.153)	(0.243)	(0.207)
Amenorrhea period	0.029	0.033	0.019	0.046
	(0.030)	(0.031)	(0.048)	(0.043)
Mother took iron tabs when pregnant	0.027	0.032	0.019	0.069
	(0.088)	(0.091)	(0.150)	(0.121)
Delivered at hospital	-0.094	-0.031	-0.033	0.004
	(0.074)	(0.078)	(0.122)	(0.105)
Delivered by caesarean	-0.115	-0.173	-0.389	0.022
	(0.167)	(0.169)	(0.253)	(0.238)
Urban	0.037	0.016	0.039	-0.000
	(0.107)	(0.122)	(0.189)	(0.171)
Imputed HIV/AIDS mothers	-0.153	-0.161	-0.191	-0.170
	(0.105)	(0.108)	(0.156)	(0.156)
Violence with partner	-0.128**	-0.154**	-0.124	-0.215**
	(0.063)	(0.068)	(0.103)	(0.095)
Mother's tribe				
Kikuyu (*Base Embu & others*)	0.311*	0.254	0.395	0.114
	(0.167)	(0.168)	(0.249)	(0.241)
Kalenjin	-0.058	-0.082	0.040	-0.227
	(0.167)	(0.171)	(0.244)	(0.248)
Kamba	-0.109	-0.113	-0.053	-0.217
	(0.170)	(0.173)	(0.247)	(0.253)
Kisii	-0.141	-0.096	-0.195	-0.045
	(0.190)	(0.200)	(0.274)	(0.307)
Luhya	-0.100	-0.082	0.077	-0.268
	(0.153)	(0.156)	(0.224)	(0.228)
Luo	-0.427***	-0.414***	-0.528**	-0.396*
	(0.158)	(0.160)	(0.229)	(0.233)
Masai	0.335	0.358	0.262	0.300
	(0.241)	(0.277)	(0.447)	(0.373)
Meru	0.276	0.212	0.115	0.193
	(0.250)	(0.251)	(0.353)	(0.368)
Mijikenda/Swahili	-0.143	-0.134	0.486	-0.547**
	(0.172)	(0.182)	(0.297)	(0.258)
Somali	-0.317*	-0.235	0.346	-0.708***
	(0.169)	(0.178)	(0.264)	(0.257)

Table 4.6: ***Continued***

	Heckman's two-step consistent estimates			
	Model 2	**Model 3**	**Model 4**	**Model 5**
Constant	0.943***	0.807***	2.022***	0.275
	(0.243)	(0.313)	(0.521)	(0.433)
	Heckman maximum likelihood estimates			
ρ	-0.232	-0.232	-0.271	-0.198
	(0.089)	(0.093)	(0.143)	(0.137)
σ	1.441	1.441	1.423	1.428
	(0.019)	(0.019)	(0.027)	(0.026)
λ (inverse Mills ratio)	-0.333	-0.334	-0.386	-0.282
	(0.130)	(0.135)	(0.207)	(0.198)
LR test of independent equations ($\rho = 0$)	$\chi^2_1 = 4.780$	$\chi^2_1 = 4.440$	$\chi^2_1 = 2.380$	$\chi^2_1 = 1.560$
	$Prob > \chi^2 = 0.029$	$Prob > \chi^2 = 0.035$	$Prob > \chi^2 = 0.123$	$Prob > \chi^2 = 0.212$

Standard errors are in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%

Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 4.7. OLS and Heckman Estimates for child's weight for age (in months)

Dependent Variable: z– score for weight-for-age					
	OLS	***Heckman's two-step consistent estimates***			
Independent Variables	**Model 1** No. obs 1,963	**Model 2** No. obs 4,075 Censored obs 363 Uncensored obs 3,712	**Model 3** No. obs 3,947 Censored obs 343 Uncensored obs 3,604	**Model 4** No. obs 1,949 Censored obs 141 Uncensored obs 1,808	**Model 5** No. obs 1,998 Censored obs 202 Uncensored obs 1,796
Child Characteristics					
Child sex (*Base - female*)	-0.202***	-0.197***	-0.198***		
	(0.056)	(0.041)	(0.041)		
Child age - all	-0.251***	-0.217***	-0.231***	-0.225***	-0.248***
	(0.033)	(0.024)	(0.025)	(0.034)	(0.036)
Child age - above 6	0.204***	0.158***	0.172***	0.171***	0.186***
	(0.039)	(0.029)	(0.029)	(0.040)	(0.042)
Child age - above 18	0.057***	0.070***	0.069***	0.063***	0.073***
	(0.014)	(0.010)	(0.010)	(0.014)	(0.014)
Child age - above 54 up to 59	-0.036	-0.030	-0.026	-0.013	-0.031
	(0.040)	(0.028)	(0.029)	(0.041)	(0.040)
Child born larger than average (*Base – largest*)	-0.031	-0.031	-0.038	-0.249*	0.193
	(0.143)	(0.098)	(0.100)	(0.144)	(0.138)
Child born average	-0.468***	-0.446***	-0.460***	-0.667***	-0.239*
	(0.133)	(0.091)	(0.093)	(0.132)	(0.129)
Child born smaller than average	-0.928***	-0.765***	-0.777***	-0.989***	-0.528***
	(0.151)	(0.106)	(0.108)	(0.151)	(0.153)
Child born very small	-0.807***	-0.877***	-0.933***	-1.068***	-0.817***
	(0.184)	(0.134)	(0.136)	(0.179)	(0.215)
Child born twins	-0.190	-0.472***	-0.456***	-0.408**	-0.526***
	(0.177)	(0.135)	(0.135)	(0.182)	(0.200)
Child received bcg vaccine	0.043	-0.086	-0.079	-0.257**	0.103
	(0.122)	(0.086)	(0.087)	(0.115)	(0.133)
Child received dpt vaccine	0.263*	0.250**	0.273***	0.262*	0.283*
	(0.154)	(0.104)	(0.106)	(0.138)	(0.165)
Child received polio vaccine	0.053	0.069	0.062	0.189	-0.030
	(0.144)	(0.099)	(0.101)	(0.138)	(0.149)
Child given vit. A last 6 months	0.023	0.021	0.017	0.024	-0.005
	(0.062)	(0.045)	(0.046)	(0.065)	(0.065)

Table 4.7: ***Continued***

	OLS		*Heckman's two-step consistent estimates*		
Independent Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Months breastfed	-0.101***	-0.095***	-0.096***	-0.099***	-0.095***
	(0.033)	(0.024)	(0.025)	(0.035)	(0.035)
Child given water	0.067	-0.031	-0.062	-0.000	-0.124
	(0.089)	(0.064)	(0.065)	(0.089)	(0.094)
Child given juice	0.050	-0.045	-0.047	0.016	-0.098
	(0.108)	(0.076)	(0.077)	(0.109)	(0.109)
Child given commercially produced baby formula	0.133	0.094	0.084	0.041	0.120
	(0.192)	(0.143)	(0.147)	(0.215)	(0.200)
Child given powdered/fresh animal milk	0.030	-0.012	-0.025	-0.089	0.031
	(0.078)	(0.056)	(0.057)	(0.079)	(0.081)
Child given pumpkin, carrots, red/yel yams, red sweet potatoes	0.033	0.059	0.040	0.053	0.057
	(0.107)	(0.075)	(0.076)	(0.107)	(0.109)
Child given green vegetables	0.042	-0.001	0.034	0.028	0.052
	(0.087)	(0.062)	(0.063)	(0.089)	(0.090)
Child given Vit. A rich fruits, eg. mango, papaya	-0.003	0.111	0.106	0.067	0.159
	(0.100)	(0.071)	(0.072)	(0.101)	(0.102)
Child given food made from local grain	-0.069	-0.012	0.011	0.002	0.018
	(0.086)	(0.063)	(0.064)	(0.090)	(0.091)
Household Characteristics					
Number of household members	-0.010	-0.018	-0.025**	-0.014	-0.028*
	(0.012)	(0.011)	(0.011)	(0.014)	(0.017)
Residence/Region					
Urban *(Base - rural)*	0.051	-0.039	-0.016	-0.122	0.076
	(0.099)	(0.075)	(0.079)	(0.110)	(0.112)
Central province *(Base - Nairobi)*	0.085	-0.050	0.008	0.012	0.062
	(0.156)	(0.110)	(0.109)	(0.150)	(0.158)
Coast province	-0.152	-0.206*	-0.065	-0.072	-0.024
	(0.155)	(0.106)	(0.121)	(0.167)	(0.173)
Eastern province	-0.206	-0.227**	-0.126	-0.138	-0.058
	(0.165)	(0.114)	(0.118)	(0.163)	(0.170)
Nyanza province	0.229	0.187*	0.258**	0.342**	0.188
	(0.155)	(0.110)	(0.117)	(0.164)	(0.167)
Rift Valley province	-0.226	-0.264**	-0.277***	-0.132	-0.355**
	(0.151)	(0.103)	(0.103)	(0.143)	(0.147)
Western province	0.040	-0.028	0.040	0.039	0.093
	(0.158)	(0.108)	(0.115)	(0.158)	(0.168)
North Eastern province	-0.131	-0.137	0.076	0.038	0.172
	(0.206)	(0.143)	(0.168)	(0.231)	(0.243)

Table 4.7: ***Continued***

	OLS		*Heckman's two-step consistent estimates*		
Independent Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Wealth level					
Poorest (*Base – richest*)	-0.306**	-0.446***			
	(0.132)	(0.096)			
Poorer	-0.217*	-0.328***			
	(0.127)	(0.094)			
Medium	-0.153	-0.234***			
	(0.125)	(0.091)			
Richer	-0.093	-0.204**			
	(0.119)	(0.087)			
Wealth, Continuous			0.671***	0.852***	0.548**
			(0.177)	(0.248)	(0.251)
Wealth Continuous residuals			-0.474**	-0.647**	-0.400
			(0.185)	(0.261)	(0.258)
Mothers Characteristics					
Education level					
With at least primary education (*Base - no education*)	0.020	0.069	0.155**	0.127	0.169*
	(0.093)	(0.070)	(0.071)	(0.099)	(0.102)
With at least secondary or higher education	0.343***	0.388***	0.400***	0.334***	0.411***
	(0.112)	(0.088)	(0.091)	(0.125)	(0.130)
Others					
HIV/AIDS positive	0.061				
	(0.110)				
Imputed HIV/AIDS mothers		0.007	-0.002	0.023	-0.061
		(0.082)	(0.083)	(0.111)	(0.124)
Births in last 5yrs	-0.093**	-0.014	-0.029	-0.017	-0.049
	(0.045)	(0.033)	(0.034)	(0.047)	(0.048)
Mother *bmi*	0.001***	0.000***	0.000***	0.001***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Violence with partner	-0.130**	-0.137***	-0.115**	-0.041	-0.194***
	(0.061)	(0.045)	(0.047)	(0.064)	(0.067)
Inverse Mills Ratio		-0.504*	-0.388	0.088	-0.358
		(0.274)	(0.280)	(0.353)	(0.346)
Constant	0.127	0.619***	0.814***	0.331	0.803**
	(0.321)	(0.231)	(0.245)	(0.358)	(0.353)

Standard errors are in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%

Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 4.8. Heckman Selection Model for child being alive or dead (for Weight)

	OLS	*Heckman's two-step consistent estimates*		
Independent Variables	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Household Characteristics				
Number of Household members	0.120***	0.124***	0.103***	0.153***
	(0.016)	(0.016)	(0.025)	(0.022)
Wealth level				
Poorest	0.042			
	(0.135)			
Poorer	0.050			
	(0.134)			
Middle	-0.031			
	(0.132)			
Richer	-0.044			
	(0.123)			
Wealth Continuous		-0.008	0.647	-0.431
		(0.276)	(0.422)	(0.387)
Wealth Continuous residuals		-0.091	-0.778*	0.383
		(0.284)	(0.439)	(0.396)
Mother's Characteristics				
Mother's education level				
With at least primary education	0.068	0.125	0.364**	-0.055
	(0.111)	(0.113)	(0.168)	(0.158)
With at least secondary or higher education	0.271*	0.281*	0.381	0.305
	(0.139)	(0.161)	(0.236)	(0.232)
Others Variables				
Age when had first sex intercourse	-0.124*	-0.127*	0.014	-0.210**
	(0.070)	(0.072)	(0.114)	(0.097)
Use physical contraception	0.183	0.153	0.150	0.171
	(0.189)	(0.189)	(0.280)	(0.273)
Use hormonal contraception	0.147*	0.147*	0.082	0.261**
	(0.077)	(0.079)	(0.118)	(0.113)
Mother's *bmi*	-0.000***	-0.000***	-0.001***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Mother given tetanus injection	0.237*	0.230	0.575**	-0.039
	(0.141)	(0.146)	(0.232)	(0.198)

Table 4.8: ***Continued***

	OLS	***Heckman's two-step consistent estimates***		
Independent Variables	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Pre-natal clinic visits (doctor/nurse/midwife)	0.348**	0.380**	0.215	0.545***
	(0.144)	(0.150)	(0.239)	(0.202)
Amenorrhea period	0.022	0.026	0.012	0.038
	(0.030)	(0.031)	(0.048)	(0.043)
Mother took iron tabs when pregnant	0.018	0.021	0.020	0.049
	(0.088)	(0.091)	(0.149)	(0.121)
Delivered at hospital	-0.088	-0.029	-0.036	0.002
	(0.074)	(0.077)	(0.121)	(0.104)
Delivered by caesarean	-0.115	-0.184	-0.408	0.009
	(0.166)	(0.168)	(0.250)	(0.237)
Urban	0.062	0.033	0.072	0.023
	(0.106)	(0.121)	(0.187)	(0.169)
Imputed HIV/AIDS mothers	-0.201*	-0.202*	-0.234	-0.232
	(0.104)	(0.107)	(0.161)	(0.149)
Violence with partner	-0.130**	-0.155**	-0.136	-0.207**
	(0.063)	(0.067)	(0.102)	(0.094)
Mother's tribe				
Kikuyu (*Base Embu & others*)	0.300*	0.241	0.395	0.091
	(0.166)	(0.167)	(0.248)	(0.240)
Kalenjin	-0.064	-0.090	0.043	-0.250
	(0.167)	(0.169)	(0.243)	(0.247)
Kamba	-0.122	-0.130	-0.063	-0.244
	(0.169)	(0.172)	(0.246)	(0.251)
Kisii	-0.148	-0.105	-0.190	-0.074
	(0.189)	(0.199)	(0.272)	(0.305)
Luhya	-0.119	-0.103	0.076	-0.320
	(0.153)	(0.155)	(0.224)	(0.227)
Luo	-0.415***	-0.409***	-0.533**	-0.387*
	(0.156)	(0.158)	(0.226)	(0.231)
Masai	0.335	0.362	0.304	0.287
	(0.240)	(0.276)	(0.446)	(0.373)
Meru	0.285	0.222	0.152	0.175
	(0.247)	(0.248)	(0.349)	(0.364)
Mijikenda/Swahili	-0.160	-0.151	0.471	-0.571**
	(0.172)	(0.181)	(0.297)	(0.256)
Somali	-0.308*	-0.225	0.353	-0.690***
	(0.169)	(0.177)	(0.263)	(0.256)
Constant	0.952***	0.835***	2.038***	0.290
	(0.242)	(0.311)	(0.515)	(0.434)

Table 4.8: ***Continued***

	OLS	***Heckman's two-step consistent estimates***		
Independent Variables	**Model 2**	**Model 3**	**Model 4**	**Model 5**
	Heckman maximum likelihood estimates			
ρ	-0.325	-0.299	0.301	-0.337
	(0.123)	(0.137)	(0.271)	(0.172)
σ	1.242	1.241	1.219	1.249
	(0.019)	(0.019)	(0.028)	(0.028)
λ (inverse Mills ratio)	-0.403	-0.372	0.367	-0.421
	(0.156)	(0.173)	(0.336)	(0.221)
LR test of independent equations ($\rho = 0$)	$\chi_1^2 = 3.700$	$\chi_1^2 = 2.550$	$\chi_1^2 = 0.450$	$\chi_1^2 = 1.860$
	$Prob > \chi^2 = 0.054$	$Prob > \chi^2 = 0.110$	$Prob > \chi^2 = 0.500$	$Prob > \chi^2 = 0.172$

Standard errors are in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%

Note: Marginal effects are for discrete change of dummy variable from 0 to 1

Table 4.9. **Tabulation of household size and child survival - for selection Model for z-score for-bmi-for-age**

Number of Household members	Child being Alive or Dead		
	0	**1**	**Proportion of children alive**
1	6	15	71.4
2	37	110	74.8
3	73	609	89.3
4	103	859	89.3
5	67	993	93.7
6	77	882	92.0
7	53	722	93.2
8	32	475	93.7
9	14	337	96.0
10	18	183	91.0
11	8	87	91.6
12	7	76	91.6
13	2	47	95.9
14	0	10	100.0
15	0	11	100.0
16	0	4	100.0
17	4	12	75.0
18	0	5	100.0
19	1	1	50.0
20	0	2	100.0
24	0	7	100.0
Total	**l,502**	**5,447**	

Tests using:

Pearson chi2(20) = 111.1262 Pr = 0.000

likelihood-ratio chi2(20) = 92.3345 Pr = 0.000

Figure 4.1a. **Child bmi-for-age z-score: Different age groups**

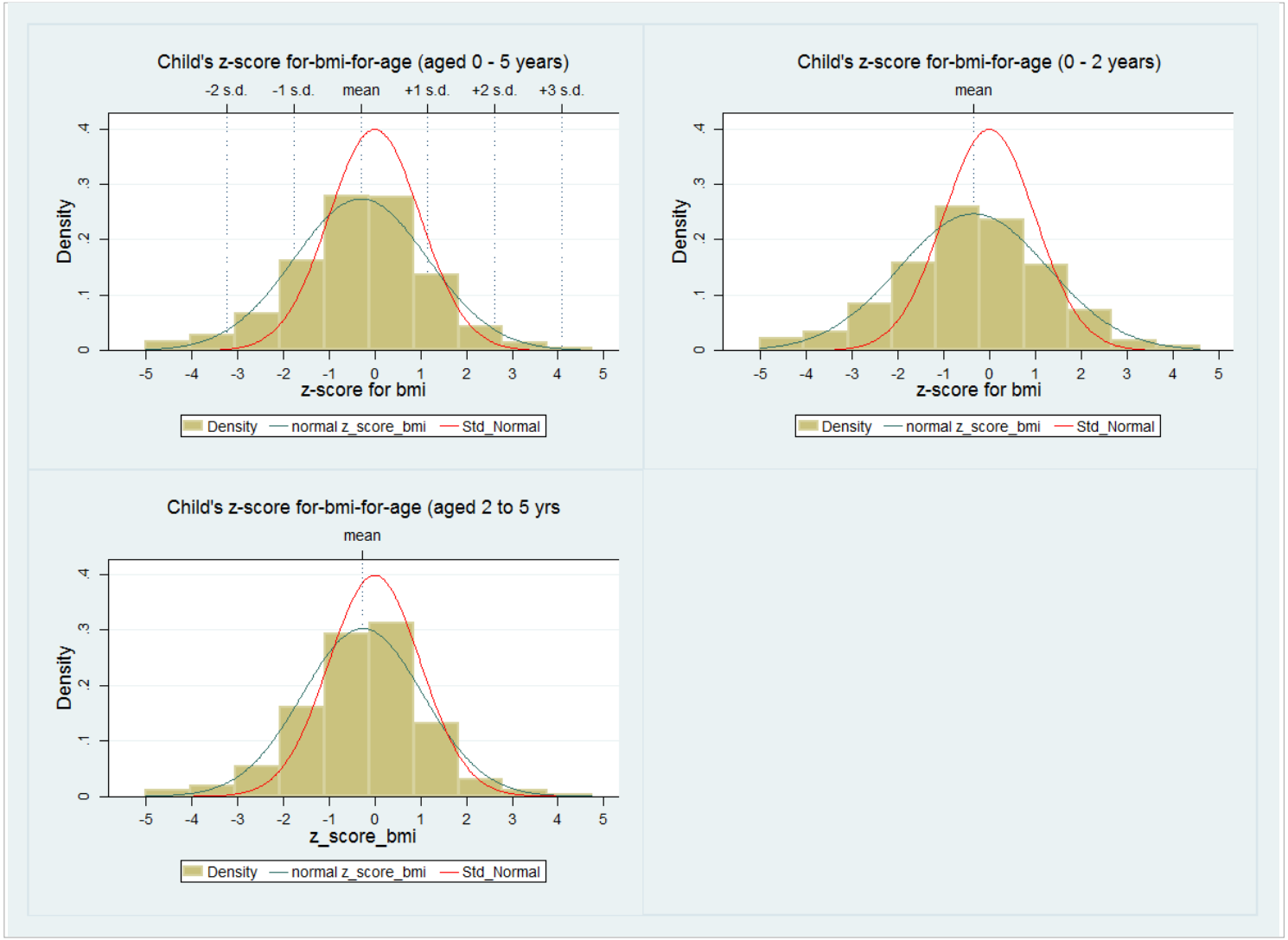


Figure 4.1b. **Child height for age z-score: Different age groups**

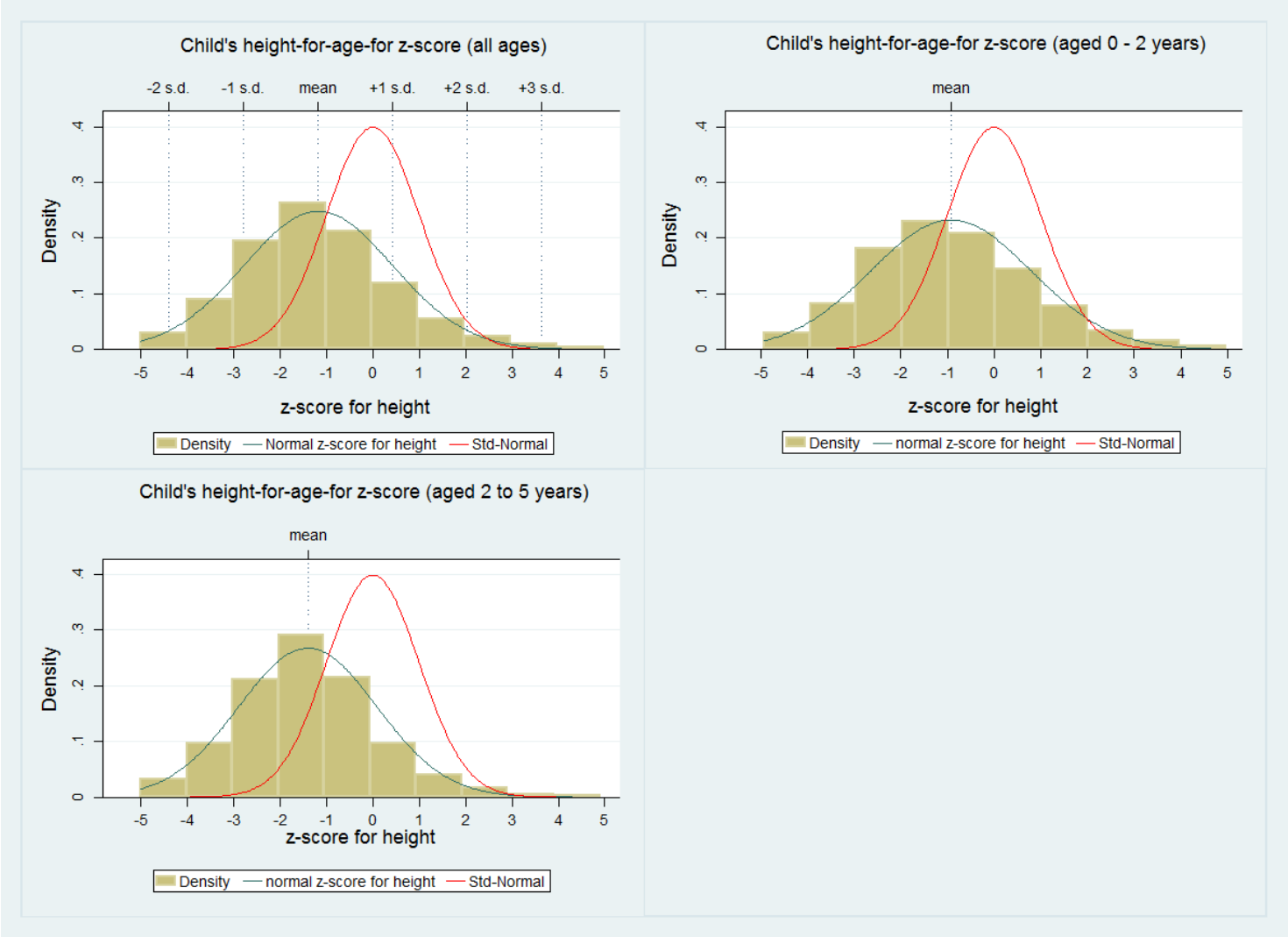


Figure 4.1c. **Child weight-for-age z-score: Different age groups**

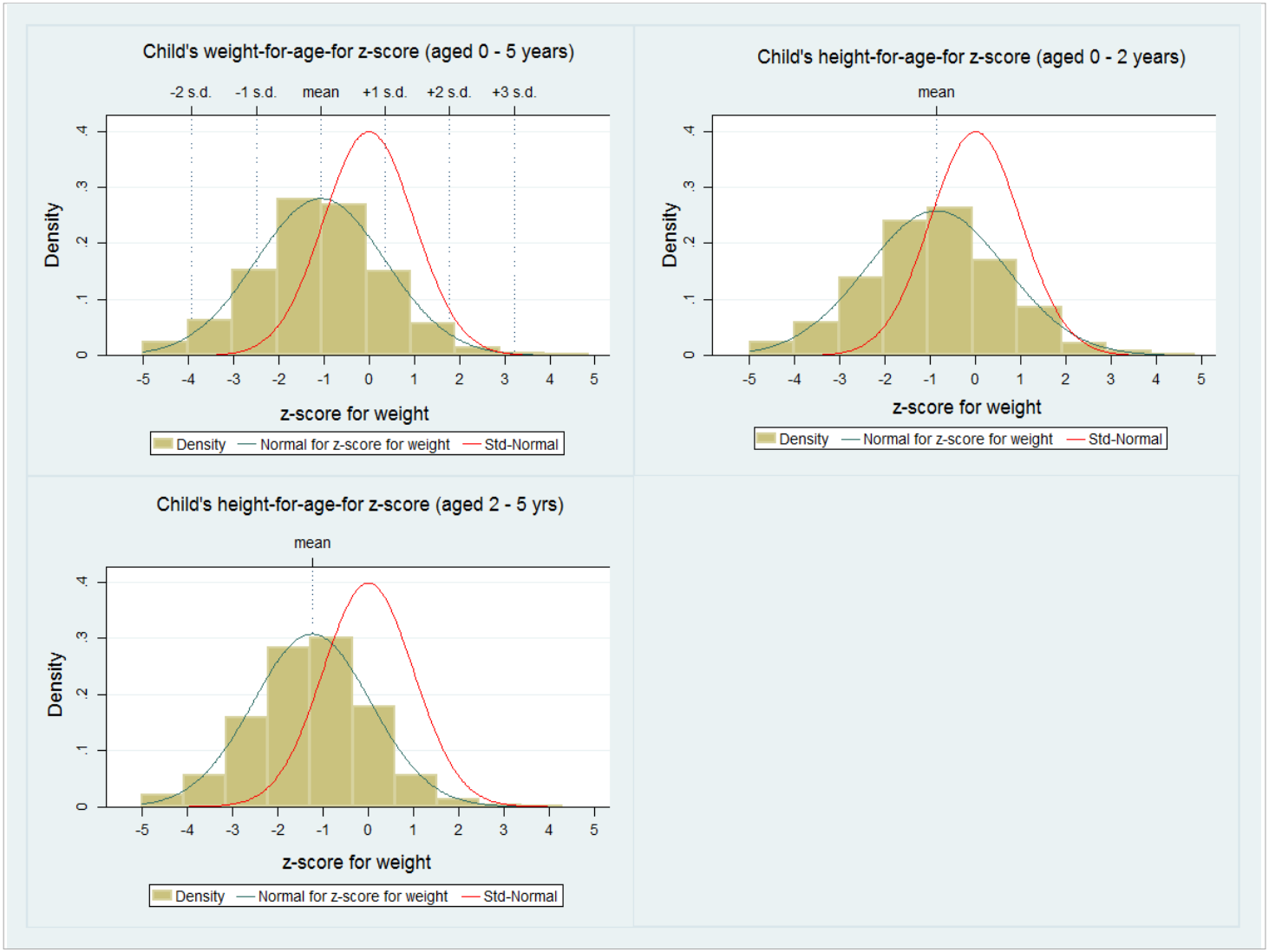


Figure 4.1d. **Scree plot of eigenvalue after PCA for household Assets**

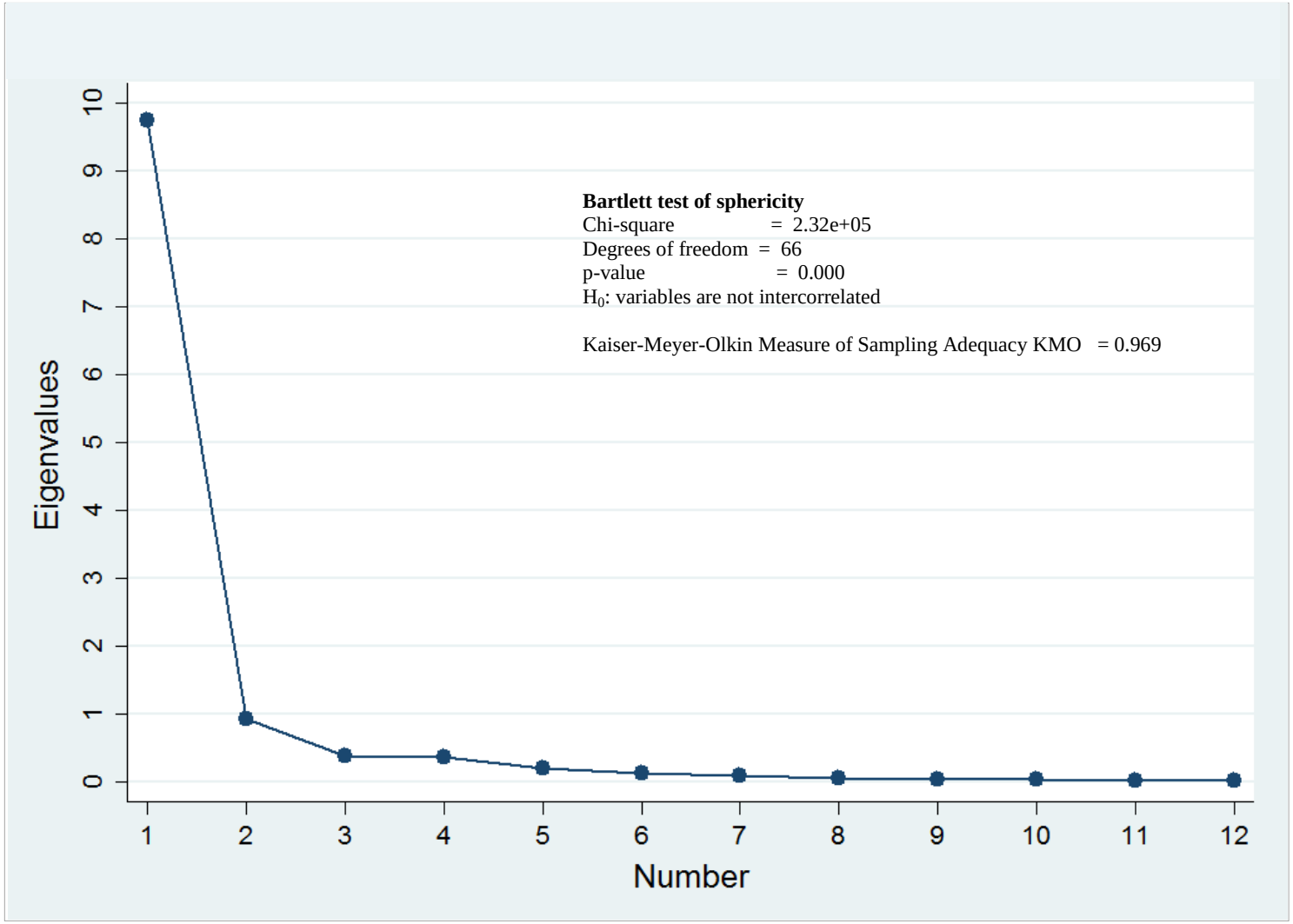
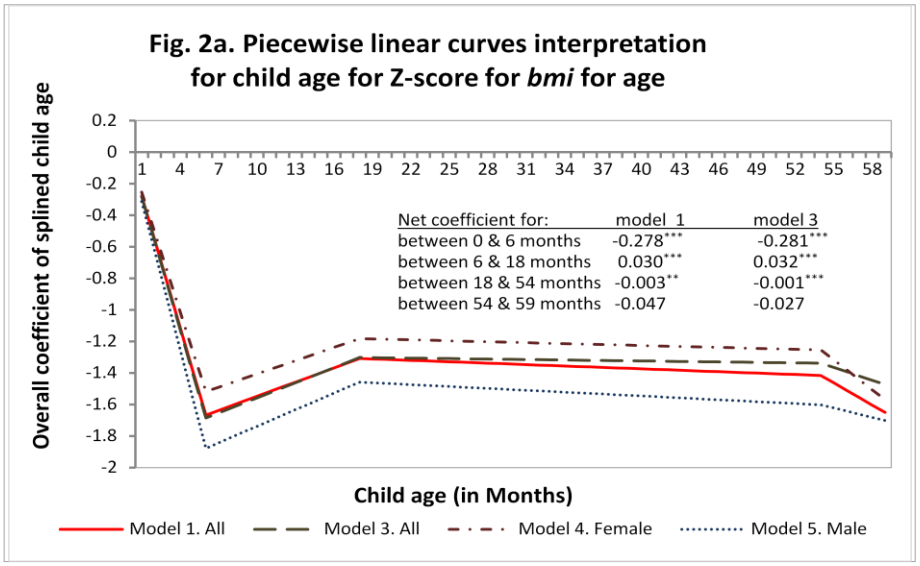
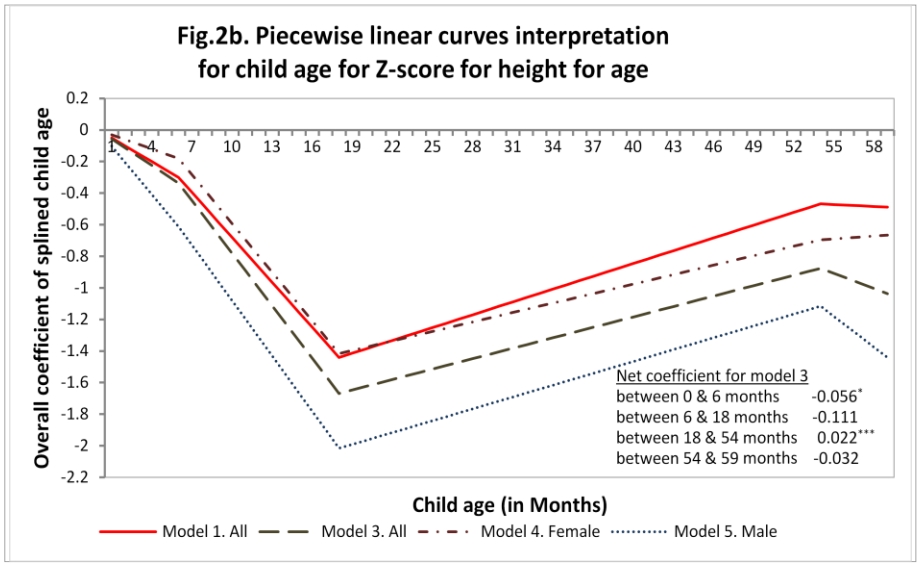
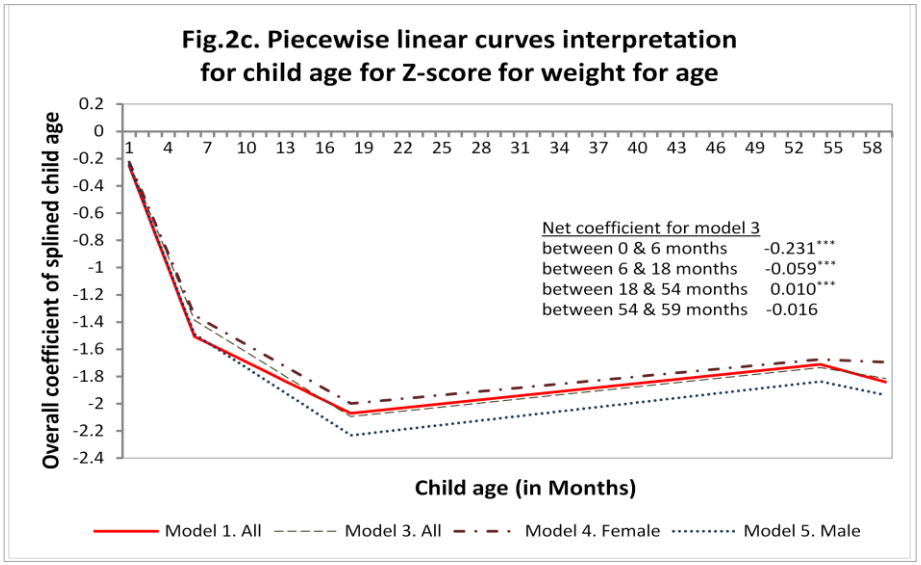


Figure 4.2. **Piecewise Linear curves interpretations**







Appendix 4.A: Variable Glossary

Dependent Variables	
z-score for body mass index (*bmi*) for age z-score for height for age z-score for weight for age	Standardized value: number of Standard Deviations a particular *bmi* or height or weight of the child is from the median value.
Independent Variables	**Description**
Household characteristics	
Household size	Dummy = 1 if household member is male and 0 if female
Polygamy	Dummy = 1 if woman is married with other wife (ves) and 0 if only wife or single.
Wealth Index	
Poorest household (*Base*) Poorer household Middle household Richer household Richest household	Dummy = 1 if the individual is in the classified household wealth named and 0 otherwise
Wealth, Continuous	Predicted score from the first principle component from PCA for household assets
Wealth Continuous residuals	Residuals after regressing Wealth (Continuous) variable on selected explanatory variables.
Residence/Region	
Urban (*Base Rural*)	Dummy = 1 if the individual is in the household residing in urban area and 0 if Rural areas
Nairobi *(Base)* Central Coast Eastern Nyanza Rift Valley Western North Eastern	Dummy = 1 if the individual is from the household residing in named the Province and 0 otherwise
Individual characteristics	
Child Characteristics	
Child sex (*Base - female*)	Dummy = 1 if child is male and 0 if female
Child age - all	Child age in months
Child age - above 6 Child age - above 18 Child age - above 54 up to 59	Child age splined at 6, 18 and 54
Child born largest (*Base*)	Dummy = 1 if child size at birth is largest
Child born larger than average	Dummy = 1 if chid size at birth is larger than average
Child born average	Dummy = 1 if child size at birth is average

Child born smaller than average	Dummy = 1 if child size at birth is smaller than average
Child born very small	Dummy = 1 if child size at birth is very small
Child born twins	Dummy = 1 if child is born in twins and 0 if born single
Child received *bcg* vaccine	Dummy = 1 if child received *bcg* vaccine and 0 if not
Child received *dpt* vaccine	Dummy = 1 if child received *dpt* vaccine and 0 if not
Child received *polio* vaccine	Dummy = 1 if child received *polio* vaccine and 0 if not
Child given Vit. A last 6 months	Dummy = 1 if child was given Vitamin A in last 6 months and 0 if not
Months breastfed	Dummy = 1 if child was breastfed and 0 if not
Child given water	Dummy = 1 if child was given water and 0 if not
Child given juice	Dummy = 1 if child was given juice and 0 if not
Child given commercially produced baby formula	Dummy = 1 if child was given commercially produced baby formula milk and 0 if not
Child given powdered/fresh animal milk	Dummy = 1 if child was given powdered or fresh animal milk and 0 if not
Child given pumpkin, carrots, red/yel yams, red sweet potatoes	Dummy = 1 if child was fed on pumpkin and/or carrots and/or red or yellow yams, and/or red sweet potatoes and 0 if not
Child given green vegetables	Dummy = 1 if child was fed on green leafy vegetables and 0 if not
Child given Vit. A rich fruits, eg. mango, papaya	Dummy = 1 if child was given Vitamin A rich fruits such as mango, papaya, etc and 0 if not
Child given food made from local grain	Dummy = 1 if child fed on food made from local grains such as millet, sorghum, etc and 0 if not
Mothers Characteristics	
Tribes	
Kalenjin	Dummy = 1 if the individual is from Kalenjin tribe and 0 if other tribe
Kamba	Dummy = 1 if the individual is from Kamba tribe and 0 if other tribe
Kikuyu	Dummy = 1 if the individual is from Kikuyu tribe and 0 if other tribe
Kisii	Dummy = 1 if the individual is from Kisii tribe and 0 if other tribe
Luhya	Dummy = 1 if the individual is from Luhya tribe and 0 if other tribe
Luo	Dummy = 1 if the individual is from Luo tribe and 0 if other tribe
Masai	Dummy = 1 if the individual is from Masai tribe and 0 if other tribe
Meru	Dummy = 1 if the individual is from Meru tribe and 0 if other tribe
Mijikenda/Swahili	Dummy = 1 if the individual is from Mijikenda/Swahili tribe and 0 if other tribe
Somali	Dummy = 1 if the individual is from Somali tribe and 0 if other tribe
Taita	Dummy = 1 if the individual is from Taita tribe and 0 if other tribe

Turkana	Dummy = 1 if the individual is from Turkana tribe and 0 if other tribe
Educational Level	
No education (*Base*)	Dummy = 1 if individual has pre Primary or no education and 0 otherwise
With at least Primary education	Dummy = 1 if individual has some Primary education and 0 otherwise
With at least Secondary or higher education	Dummy = 1 if individual has Secondary or higher education level and 0 otherwise
Other mothers Characteristics	
HIV/AIDS positive	Dummy = 1 if individual was HIV/AIDS positive and 0 if negative
Imputed HIV/AIDS mothers	Dummy = 1 if imputed individual was HIV/AIDS positive and 0 if negative
Births in last 5yrs	Number of births mother had in previous last 5 years
Mother *bmi*	Mothers *bmi*
Violence with partner	Dummy = 1if there was violence in the household and 0 if no violence
Instruments for selection model	
Age when had first sex intercourse	Dummy = 1 if mother had sexual intercourse under the age of 15 years and 0 otherwise
Use physical contraception (*base = if use no contraception*)	Dummy = 1 if mother used physical contraception to prevent pregnancy and 0 otherwise
Use hormonal contraception	Dummy = 1 if mother used hormonal contraception to prevent pregnancy and 0 otherwise
Mother given tetanus injection	Dummy = 1 if mother received tetanus injection during pregnancy, before delivery and 0 otherwise
Pre-natal clinic visits (doctor/nurse/midwife)	Dummy = 1 if mother visited clinics and seen by doctors or nurse or midwife during her pregnancy period and 0 otherwise
Amenorrhea period	Dummy = 1 if mother had any amenorrhea period and 0 otherwise
Mother took iron tabs when pregnant	Dummy = 1 if mother took iron tablets during pregnancy period and 0 otherwise
Delivered at hospital	Dummy = 1 if mother delivered at the hospital or clinic and 0 otherwise
Delivered by caesarean	Dummy = 1 if mother had child through caesarean and 0 if normal
λ	Inverse Mills Ratio: Selectivity bias correction factor computed from the estimated household member testing HIV/AIDS positive

CHAPTER 5

CONCLUSION

There are three major studies in this thesis examining issues of individual HIV/AIDS status, children's education and health of young children.

Chapter 2 of this thesis considered the factors which contribute to a healthy individual acquiring the HIV disease. Specifically, the study examined the impact of household and individual characteristics on the likelihood of an individual adult acquiring the HIV disease. The main technical issues surrounded the endogeneity of household wealth and non-random sample selection brought about by how HIV testing was done in the sample. Principal component analysis was used to construct a continuous wealth variable, and a Rivers Voung procedure was used to correct for endogeneity of household wealth. A Heckman procedure was used to solve the problem of non-random selection of the sample. Although three models were examined, the conclusions of the study are based on the third model which addresses these two problems, giving consistent and reliable estimates. The study found several factors with substantial contribution to the spread of HIV in Kenya. These factors include: gender of the household head, wealth, individual occupations, marital status, social behaviour such as multiple sex partners and use of condoms, cultural practices such as polygamy due to traditional cultures and customs around the treatment of women. Even after controlling for these factors, there are quite sizeable differences in the likelihood of being HIV positive by geographical and ethnic tribe. This suggests the need for new policies and projects to focus on and address these factors which are deeply embedded in cultural / tribal practices in certain parts of the

country. Such projects could include education programs to the most affected areas, especially in the rural areas to provide the communities with new understanding and raise awareness of the risks created by their traditional, cultural and other factors in question towards the contraction and spread of HIV/AIDS. In addition, these HIV programs must address the root causes of gender-based vulnerability to HIV. Governments must adopt policies and enact legislation against harmful traditional practices that increase vulnerability to HIV, including the violence against women and sexual minorities. It is suggested that measures to be taken to introduce sex education curriculum in schools, and boys and girls should be provided with information on HIV prevention. Other bodies including leaders and religious leaders could be encouraged to engage in teaching of behaviour change interventions such as promoting condom use.

The third chapter examines child schooling outcomes, specifically school attendance, grade attainment and grade progression rate. To assess a child's current classroom engagement, a probit model is used to model attendance at primary school and secondary school. The school attainment model captures the likelihood of children completing primary education, with use of an ordered probit model. The rate of grade progression is defined as the ratio of the number of grades completed to the number that the child should have completed given their age, had they progressed one grade each year. In each of these models of schooling outcomes several different specifications were used. These deal with the availability of wealth as only a categorical variable, and correct for the endogeneity of wealth. Modelling also has to account for the non-random selection of the sample when a variable for HIV status is included, as this includes many missing observables. To avoid the loss of sample size due to missing observations on the HIV data, we also show results where the missing data has been inferred by imputation

using chain equations. These models examine the impact of household characteristics, socio-cultural factor and HIV/AIDS on these three child schooling outcomes

The study found impact of several factors on a child's school outcomes. Although male children were generally found to attend secondary school more than females, a higher percentage of female children finished primary school and with higher rates of grade progression for age than males. The study found some rivalry effects as well synergy effects among the children in the same household. The presence of young pre-primary school aged children tends to have a detrimental effect on the schooling of older girls, while there are some benefits to having older siblings at a similar stage of schooling. The loss of parents (either the mother or the father) has a detrimental effect on both school attendance and rates of grade progression. Those from the Islamic religion also have significantly lower attendance and rates of grade progression. The strongest effect was, not surprisingly, found with the education level of the household head, which has a strong positive effect on schooling. If there is an individual in the household with HIV/AIDS, this was found to detrimentally affect a child's secondary school attendance and rates of grade progression especially in the urban areas. Household wealth is also a key factor influencing educational attainment.

Chapter four investigates the factors influencing the health of young children in Kenya. The factors used in the models included the child's characteristics, household and mother's characteristics. To measure child health, the study uses three indicators: body mass index (*bmi*) for age, height for age and weight for age. Z-scores were obtained to standardise these indicators to allow them to be compared across age and sex. This transformation enabled easy interpretations of the results because the scale becomes linear and standardised. The transformation used UK standards and compared to Kenyan

measurements the z-scores appeared approximately normally distributed with the mean values being negative for the three health indicators.

A number of similar econometric issues were encountered as in Chapter 3, including the endogeneity of wealth, sample selection bias, and missing HIV data as well as sample selection bias due to the fact that not all children survive. Similar techniques were adopted to deal with these issues. The estimation results indicate several factors affect a child's health during their early years. The results established the height for age and weight for age of Kenyan children are inferior to those given in the UK standards, especially for males. A number of maternal characteristics play a critical role in child health, and the child's pre-birth and birth experiences are critical. After birth, nutrient intake plays a major role. Wealth has a strong effect on long term health status, as measured by height for age, but little impact on short term health episodes. Mother's characteristics including education and her health (*bmi*) were found to improve a child's health. However, other issues in the household such as domestic violence and maternal HIV/AIDS do not appear to have a detrimental effect on children's health.

Although the studies in this thesis do not deal directly with policy and strategic issues, our findings gives important insight which can be very valuable to both researchers and policy makers. These results open a gateway for further research work in the fields of HIV/AIDS, child schooling and health issues in Kenya. The more recent data sets which are available could be used together with previous data sets to provide a comprehensive picture. This will help examine the impact of HIV/AIDS on children and households. An evaluation of the impact of HIV/AIDS on economic development of the households and the country as a whole could give insight which would help strategising and improving the policies and programs to reduce the prevalence and spread of the disease.

In 2003 the Kenyan government introduced and implemented a Free Primary Education program (FPE). The program has faced major challenges including lack of facilities, few teachers, lack of finances to support the program, over-age children enrolments. There is a critical need to examine child schooling and evaluation of the program to enable improvement of the schooling outcome and education system in Kenya. Results in this thesis suggest progression though grades is a major issue, indicative of constraints in the supply of quality schooling, as well as the impact of household commitments (household chores, caring for younger siblings, etc) on children's ability to engage with schooling. It will take a lot more than the provision of free primary education to see significant improvement in educational outcomes.

In recent years, there has been a significant improvement of health facilities and child health services in Kenya. At the same time, the results in this thesis show there is a long way to go. Access to health services, especially secondary and tertiary, is still quite inequitable, and many of the broader issues of public health, including nutrition and sanitation, are clearly a challenge, given the high percentage of relatively poor health outcomes for young children especially.

As with any cross sectional data set, the data used in this thesis relies on the differences between the subjects of research interest, but does not show us changes over time. In addition, there are omitted important variables for the modelling which might have given a further understanding of factors affecting the questions in the study. The sampling and the survey data obtained could lead to inappropriate use of econometric methods and tools. Although the study tries to use the best available econometric tools, the limitations of this survey data cause us to interpret the findings with caution until further research can be undertaken.

BIBLIOGRAPHY

Abagi O. and G. Odipo (1997). Efficiency of Primary Education in Kenya: Situational Analysis and Implications for Educational Reform. Discussion Paper DP004/97, Institute of Policy Analysis and Research.

Abagi, O. (1997a). Efficiency of primary education in Kenya: situational analysis and implications for educational reforms. *IPAR Discussion Paper No. 4. Nairobi: Institute of Policy Analysis and Research.*

Abagi, O. (1997b). Status of education in Kenya: indicators for planning and policy formulation. *IPAR Special Report. Nairobi*: *Institute of Policy Analysis and Research.*

Abagi, O., W. Owino, D. N. Sifuna, M. Waga, C. Ngome, D. Aduda, and A. Karugu (2000). Implementing The Report Of The Commission Of Inquiry Into The Education System Of Kenya (The Koech Report): Realities, Challenges And Prospects. *IPAR special report series, SR NO. 03/2000.*

Abate M. (1993). Factors responsible for childhood mortality variation in rural Ethiopia *Journal of Biosocial Science.* Vol. 25: 223-238.

About Kenya (2004). Retrieved 15 Nov. 2010 , from *http://www.jambokenya.com/jambo/kenya/kenya.htm*

Acemoglu, D. and J. Angrist (2001). How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws, NBER Macroeconomics Annual 2000, *National Bureau of Economic Research, Inc*. Vol. 15: 9-74.

Adams S. (2006). The impact HIV/AIDS has had on unemployment and the contributing factors. SA *Pharmaceutical Journal.* Vol. 73(5):46-47.

African Studies Center [ASC] - *Kenya Ethnic Groups.* Retrieved 25 August 2008, from *http://www.africa.upenn.edu/NEH/kethnic.htm*

Ainsworth, M., K. Beegle and G. Koda (2002). The impact of adult mortality on primary school enrolment in north-western Tanzania. *Human Development Sector: Working Paper. Africa Region, The World Bank.*

Ahmed S, T. Lutalo, M. Wawer, *et al*. (2001). HIV incidence and sexually transmitted disease prevalence associated with condom use: a population study in Rakai, Uganda. *AIDS.* Vol.15(16):2171-2179.

Akhtar N., M. Iqbal Zafar, S. Nighat, A. Safdar, S. Rasheed, S. Saddique and A. Kousar (2005). Factors Affecting Child Health: A Study of Rural Faisalabad. *Journal of Agriculture and Social Sciences*. Vol.1(1):50–53.

Akresh, R. (2004). Adjusting Household Structure: School Enrolment Impacts of Child Fostering in Burkina Faso. *IZA* Working Paper # 1379.

Allison, P. D. (2001). Missing Data, *Sage University Papers Series on Quantitative Applications in the Social Sciences*, series no. 07-136, Thousand Oaks: Sage Publications

Alvarezm R. M. and G. Glasgow (2000). Two – Stage Estimation of Norecursive Choice Models. *Political Analysis* Vol.8(2):11-24.

Ambasa-Shisanya C. (2007). Widowhood in the era of HIV/AIDS: A case study of Slaya District, Kenya. *Journal of Social Aspects of HIV/AIDS*. Vol.4(2): 606-614.

Ambasa-Shisanya C. R. (2007). Widowhood in the era of HIV/AIDS: a case study of Siaya District, Kenya. *Sahara Journal*. Vol.4(2):606-15.

Amemiya, T. (1978). The Estimation of a Simultaneous Equation Generalized Probit Model, *Econometrica,* Vol.46(1):193-205.

Amnesty International Document (2002). Kenya: Rape - The Invisible Crime. AFR 32/001/2002

Anderson, K. G. (2005). Relatedness and investment in children in South Africa. *Human Nature* Vol.16(1):1-31.

Anderson, K. G., and D. Lam, (2003). Dynamics of family structure and progress through school in South Africa: evidence from retrospective histories. *Paper presented at the Population Association of America annual meeting*, Minneapolis, MN.

Anh, T. S., J. Knodel, D. Lam, and J. Friedman (1998). Family Size and Children's Education in Vietnam. *Demography*, Vol.35(1):57-70.

Arendt J. N. and A. Holm (2006). Probit models with binary endogenous regressors, *Discussion Papers on Business and Economic*, University of Southern Denmark.

Arrow, K. J. (1973). Higher education as a filter. *Journal of Public Economics*, Vol.2(3):193-216.

Australian Government – AusAID (2009). HIV/AIDS – A Global Problem. Retrieved 10th Sept. 2010, from *http://www.ausaid.gov.au/closeup/hivaids.cfm*

Ayikukwei, R. M., D. Ngure, J. E. Sidle, D. O. Ayuku, J. Baliddawa and J. Y. Greene (2007). Social and Cultural Significance of the Sexual Cleansing Ritual and its Impact on HIV Prevention Strategies in Western Kenya. *Sexuality and Culture*. Vol.13(3):32-50.

Bailes E., F. Gao, *et al.* (2003) Hybrid Origin of SIV in Chimpanzees, *Science.* Vol. 300(5626):1713

Baker, L. L., P. G. Jaffe, L. Ashbourne and J. Carter. (2002). Children Exposed to Domestic Violence: A Teacher's Handbook to Increase Understanding and Improve Community Responses. Centre for Children and families in the justice system. *National Education Association*, London, Ontario, Canada

Bakis O., N. Davutyan, H. Levent, and S. Polat (2009). Social returns to higher education in Turkey: a quantile regression approach. *Shock Vulnerability and Therapy, Economic Research Forum 16th Annual Conference*. Cairo, Egypt.

Barman S. (2010). Parental Education, Parental Death, Poverty and Socio-economic Impact on School Attendance Status of Children in India. *Academic Leadership journal*. Vol 8(4).

Barnighausen T, V. Hosegood, I. M. Timaeus, and M. L. Newell (2007). The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. *AIDS*. Vol.21(7):29–38.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Statistical Society Series A, Mathematical and Physical Sciences.* Vol.160(901):268-282.

BBC News. (2003). Is tradition to blame for spreading HIV? Retrieved May, 2009, from *http://news.bbc.co.uk/2/hi/africa/3275539.stm*

Becker, G. S. (1964). Human Capital, 1st ed. *New York: Columbia University Press for the National Bureau of Economic Research*.

Becker, G. S. (1965). A Theory of the Allocation of Time. *Economic Journal*, Vol. 75: 493-517.

Becker, G. S. (1967). Human Capital and the Personal Distribution of Income: An Analytical Approach. *Institute of Public Administration.* University of Michigan, Ann Arbor, Michigan. pp 49.

Becker, G. S. (1975). Human Capital: A Theoretical and Empirical Analysis With Special Reference to Education. Second Edition. *National Bureau of Economic Research Inc.*, beck75-1, New York.

Becker, G. S. (1994). Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition), *NBER Books, National Bureau of Economic Research, Inc*, number beck94-1, New York.

Becker, G. S., and H. G. Lewis (1973). Interaction between Quantity and Quality of Children. *Journal of Political Economy,* Vol.81(2), 279 - 288.

Bedi A.S., P. K. Kimalu, D. K. Manda, and N. N. Nafula (2004). The Decline in Primary School Enrolment in Kenya. *Journal of African Economies,* Vol.13(1): 1-43.

Benhabib, J. and M.M. Spiegel (1994). The role of human capital in economic development. Evidence from aggregate cross-country data. *Journal of Monetary Economics.* Vol.34(2): 143-173.

Bennell, P., B. Chilisa, K. Hyde, A. Makgothi, E. Molobe and L. Mpotokwane (2001). The impact of HIV/AIDS on Primary and Secondary Education in Botswana: Developing a Comprehensive Strategic Response. *Ministry of Education, Government of Botswana Department for International Development*, United Kingdom

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy,* Vol.75(1):352-65.

Berry, N. C. Davis, A. Jenkins, D. Wood, P. Minor, G. Schild, M. Bottiger, H. Holmes and N. Almond (2001), Vaccine safety: Analysis of oral polio vaccine CHAT stocks. *Nature,* Vol.410:1046-4047.

Blancou, P. Jean-Pierre Vartanian, C. Christopherson, N. Chenciner, C. Basilico, S. Kwok and S. Wain-Hobson (2001). Polio vaccine samples not linked to AIDS. *Nature,* Vol.410:1045-1046.

Bloom S. S., M. Urassa, J. Ng'weshemi, and J. T. Boerma (2002). Community effects on the risk of HIV infection in rural Tanzania. *Sexual Transmitted Infections*. Vol. 78(4):261–266.

Boerma J. T., S. Gregson, C. Nyamukapa, and M. Urassa. (2003). Understanding the uneven spread of HIV within Africa: comparative study of biologic, behavioral, and contextual factors in rural populations in Tanzania and Zimbabwe. *Sexually Transmitted Diseases*. Vol.30(10):779–787.

Bogonko, S. N. (1992). A History of Modern Education in Kenya (1895 – 1991), *Evans Brothers (Kenya),* Nairobi, Kenya. Pp. 208.

Bogonko, S. N. (1992). Reflections on Education in East Africa. Nairobi: *Oxford University Press*. Pp 326.

Bound, J. Jaeger, D. A. and Barker, R. M (1995). Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak. *Journal of the American Statistical Association*, Vol. 90(430):443-450.

Bowen P., R. Dorrington, G. Distiller, H. Lake and S. Besesa, (2008). HIV/AIDS in the South African construction industry: an empirical study. *Construction Management and Economics, Taylor and Francis Journals*, Vol.26(8): 827-839.

Boyle, E. H. (2002). Female Genital Cutting: Cultural Conflict in the Global Community. Baltimore: *Johns Hopkins University Press.* Pp.188.

Brass, W. (1993). Child mortality improvement and the initiation of fertility falls in Kenya, International Union for the Scientific Study of Population (IUSSP) (ed.), *International Population Conference, Montreal*, Vol. 1:73-79. IUSSP, Liege.

Bureau of African Affairs. *U.S Department of State*. Retrieved 25 Feb. 2011, from *http://www.state.gov/r/pa/ei/bgn/2962.htm*

Butcher, K. F., and A. Case, (1994). The Effect of Sibling Sex Composition on Women's Education and Earnings. *Quarterly Journal of Economics*, Vol. 109(3), 531- 563.

Buve A., K. Bishikwabo-Nsarhaza, and G. Mutangadura. (2002). The spread and effect of HIV-1 infection in sub-Saharan Africa. *Lancet*. Vol.359:2011-7.

Bwayo, J. J., F. Plummer, M. Omari, A. Mutere, S. Moses, J. Ndinya-Achola, P. Velentgas and J. Kreiss, (1994). Human immunodeficiency virus infection in long-distance truck drivers in East Africa. *Archives of Internal Medicine.* Vol. 154: 1391-1396.

Bwayo, J.J., A.N. Mutere, M.A. Omari, J.K. Kreiss, W. Jaoko, C. Sekkade- Kigondu and F.A. Plummer. (1991). Long distance truck drivers 2: Knowledge and attitudes concerning sexually transmitted diseases and sexual behaviour. *East African Medical Journal.* Vol. 68(9): 714-719.

Caldwell J. C. (1979). Education as a Factor in Mortality Decline: An Examination of Nigerian Data. Population Studies, Vol. 33(3): 395-413.

Caldwell J.C. (1994). How is greater maternal education translated into lower child mortality? *Health Transition Review.* Vol. 4: 224–29.

Card, D. (1999). The causal effect of education on earnings. Handbook of Labor Economics, in: O. Ashenfelter & D. Card (ed.), Handbook of Labor Economics, ed 1, Vol. 3(3): chapter 30: pp 1801-1863. Elsevier.

Carson A. J., R. Sandler, F. N. Owino, F. O. Matete, and E. C. Johnstone, (1998). Psychological morbidity and HIV in Kenya. *Acta Psychiatrica Scandinavica*. Vol. 97(4):267-71.

Cascio, E. (2005). School Progression and the Grade Distribution of Students: Evidence from the Current Population Survey. IZA Discussion Paper 1747. *Institute for the Study of Labor (IZA)*. Bonn, Germany: Forschungsinstitut zur Zukunft der Arbeit.

Case A. and C. Ardington (2006). The Impact of Parental Death on School Outcomes: Longitudinal Evidence from South Africa *Demography*. Vol. 43(3): 401-420.

Case, A., C. Paxson, and J. Ableidinger (2004). Orphans in Africa: parent death, poverty, and school enrolment. *Demography* Vol.41(3): 483-508.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research.* Vol.1(2):245-276.

Centers for Disease Control and Prevention [CDC], (2008). HIV Prevalence Estimates - United States, 2006 Morbidity and Mortality Weekly Report. *Centers for Disease Control and Prevention.* 1600 Clifton Rd, MailStop E-90, Atlanta, GA 30333, U.S.A.

Centre for Maternal and Child Enquiries (CMACE) (2011). Perinatal Mortality 2009 report: United Kingdom. CMACE: London.

Central Bureau of Statistics [CBS] (Kenya): Ministry of Health (MOH) [Kenya], and ORC Macro. 2004. *Kenya Demographic and Health Survey 2003.* Calverton, Maryland: CBS, MOH, and ORC Macro.

Central Bureau of Statistics [CBS] (Kenya): Ministry of Planning and National Development. (2003a). *Economic Survey*. Nairobi, Kenya: CBS.

Central Bureau of Statistics [CBS] (Kenya): Ministry of Planning and National Development. (2003a). *Economic Survey*. Nairobi, Kenya: CBS.

Centre for Rights, Education and Awareness [CREAW] (2008). Wife Inheritance: A Death Sentence Behind the Mask of Culture. *A publication of Centre for Rights, Education and Awareness (CREAW)*. Nairobi, Kenya.

Chambers, R. (2003). Evaluation Criteria for Statistical Editing and Imputation, *National Statistics Methodological Series No* 28. Office for National Statistics, London, UK.

Chant, S. (2003). Female Household Headship and the Feminisation of Poverty: Facts, Fictions and Forward Strategies, New Working Paper Series, Issue 9, *Gender Institute, London School of Economics and Political Science*, London, UK.

Checchi D. (2006). The Economics of Education. *Cambridge University Press*, UK.

Cheluget, B., G. Baltazar, P. Orege, M. Ibrahim, L. H. Marum and J. Stover (2006). Evidence for population level declines in adult HIV prevalence in Kenya. *Sex Transm Infec.* Vol.82(1): 21-26.

Chen Y. and H. Li (2009). Mother's education and child health: is there a nurturing effect?. *Journal of health economics*. Vol.28(2):413-426

Chernichovsky, D. (1985). Socioeconomic and Demographic Aspects of School Enrolment and Attendance in Rural Botswana. *Economic Development and Cultural Change*, Vol.33(2):319-332.

Chieni, S. N. "The Harambee Movement in Kenya the Role played by Kenyans and the Government in the Provision of Education and other Social Services." *Moi University*, Kenya. Last updated 26/04/1999. Retrieved 16th Feb. 2011. From *http://boleswa97.tripod.com/chieni.htm*

Chitnis, A, D. Rawls, J. Moore. (2000). Origin of HIV Type 1 in Colonial French Equatorial Africa? *AIDS Research and Human Retroviruses*, Vol.16(1):5-8.

Claibourn, M.P., and P.S. Martin (2000). Trusting and Joining: An Empirical Test of the Reciprocal Nature of Social Capital. *Political Behavior*, Vol.22(4):267-291.

Cohen, J. (2000). The Hunt for the Origin of AIDS. *The Atlantic.* Retrieved 23rd Oct, 2011, from *http://www.theatlantic.com/magazine/archive/2000/10/the-hunt-for-the-origin-of-aids/6490/.*

Cole T. J. (1990). The LMS method for constructing normalized growth standards. *European Journal of Clinical Nutrition.* Vol. 44(1):45-60.

Cole T. J. and P. J. Green (1992). Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine* Vol.11(10):1305-19.

Cole T. J., J. V. Freeman and M. A. Preece (1998). British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Statistics in Medicine.* Vol.17(4):407-29.

Cole T. J., M. C. Bellizzi, K. M. Flegal and W. H. Dietz (2000). Establishing a standard definition for child overweight and obesity worldwide: international survey. *British Medical Journal.* Vol. 320(7244):1240–1243.

Conley, D. and R. Glauber (2005). Parental Educational Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order From Exogenous Variation in Fertility. NBER Working Paper 11302. *National Bureau of Economic Research, Inc.* New York University.

Cornwell, K., B. Inder, P. Maitra, and A. Rammohan (2005). Household Composition and Schooling of Rural South African Children: Sibling Synergy and Migrant Effects, Monash Economics Working Paper No. 11/05, *Monash University*, Australia.

Cummings, E. M. and P. D. Davies (1994). Children and marital conflict: The impact of family dispute and resolution, *Guildford Press*, New York. Pp. 194.

Counties of Kenya. Retrieved 23rd Oct. 2011, from *http://en.wikipedia.org/wiki/Counties_of_Kenya.*

Dahlgren G. and M. Whitehead (1995). Tackling inequalities: a review of policy initiatives. In: Benzeval M, Judge K, Whitehead M, editors. Tackling inequality in health: an agenda for action. *London: Kings Fund Institute*, UK.

Day P., J. Pearce and D. Dorling (2008). Twelve wolrds: a geodemographic comparison of global inequality in mortality. *Journal of Epidemiol Community Health.* Vo. 62(11):1002-1010.

De Cock K. M., M. G. Fowler, E. Mercier, I. de Vincenzi, J. Saba, E. Hoff, D. j. Alnwick, M. Rogers and N. Shaffer (2000). Prevention of mother-to-child HIV transmission in resource-poor countries–translating research into policy and practice. *JAMA:* Vol. 283 (9): 1175–1182.

Deaton A. (1997). The Analysis of Household Surveys: A Microeconometric Approach to Development Policy, Baltimore, *Johns Hopkins University Press* for the World Bank. Pp. 479.

Dehejia, R. and R. Gatti (2002). Child labour: The Role of income variability and access to credit across countries. NBER Working paper No. 9018, *National Bureau of Economic Research, Inc.* Cambridge, MA.

Desai S. and S. Alva (1998). Maternal education and child health: Is there a strong causal relationship? *Demography*. Vol.35(1):71–81.

Desai, S. (1992). Children at Risk: The Role of Family Structure in Latin America and West Africa. *Population and Development Review.* Vol.18(4):689-717.

Deschamps, M.M, J.W. Pape, A. Hafner, W.D. Johnson Jr. WD (1996). Heterosexual transmission of HIV in Haiti. *Annals of Internal Medicine*. Vol.125(4):324-330.

De Waal, A., (2006). AIDS and Power: Why There Is No Political Crisis – Yet. Zed Books Ltd. "in association with the International African Institute and the Royal African Society". New York.

De Walgue, D. (2006). Who gets AIDS and How? The determinants of HIV Infection and sexual behaviours in Burkina Faso, Cameroon, Ghana, Kenya and Tanzania. World Bank Policy Research Working Paper 3844, World Bank, Washington.

Dinkelman T, D. Lam and M. Leibbrandt, (2007). Household and community income, economic shocks and risky sexual behavior of young adults: evidence from the Cape Area Panel Study 2002 and 2005. *AIDS*. Vol. 21(7):49–56.

Dreze, J. and G. G. Kingdon (2001). School Participation in Rural India, Re*view of Development Economics,* Vol.5(1): 1-24.

Duncan, W.A. (1989). Engendering school learning: science attitudes and achievement among girls and boys in Botswana. Studies in comparative and international education, No. 16. *Institute of International Education, University of Stockholm.* Stockholm, Sweden. Pp.222.

Dziuban C. D. and E. C. Shirkey (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin.* Vol. 81(6): 358-361.

Edleson, J.L., (1999) Problems Associated with Children's Witnessing of Domestic Violence. Retrieved 10th January 2011. *http://new.vawnet.org/category/Main_Doc.php?docid=392*

Eide, E. R. and M. H. Showalter, (2001). The Effect of Grade Retention on Educational and Labor Market Outcomes. *Economics of Education Review.* Vol.20(6):563-576.

Eisemon, T.O. (1988). Benefiting from Basic Education, School Quality and Functional literacy in Kenya, 1st ed. *Pergamon Press,* Oxford, New York. Pp. 149.

Eisemon, T.O. and A. Nyamete (1988). Schooling and Agricultural Productivity in Western Kenya. *Journal of East African Research and Development, Vol.*18: 44-66.

Ellis, F. (2000) Rural Livelihoods and Diversity in Developing Countries. *Oxford University Press,* Oxford, New York. Pp.279.

Eshiwani, S.G. (1993). Education in Kenya Since Independence. *East African Educational Publishers Ltd*. Nairobi, Kenya. Pp.229.

Faber M., J. Paul, van Jaarsveld and R. Laubscher (2007). The contribution of dark-green leafy vegetables to total micronutrient intake of two- to five-year-old children in a rural setting. *Water Research Commission SA*, Vol. 33(3) (Special Edition).

Fears, D. (2005) Study: Many Blacks Cite AIDS Conspiracy, *The Washington Post. Page A02.* Retrieved on 23rd, Oct. 2011, from *http://www.washingtonpost.com/wp-dyn/articles/A33695-2005Jan24.html*.

Field C. J. (2005). The Immunological Components of Human Milk and Their Effect on Immune Development in Infants. *The American Society for Nutritional Sciences*. Vol. 135(1):1-4.

Filmer D. and L. Pritchett (1999). The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries, *Population and Development Review*. Vol.25(1), 85 – 120.

Filmer, D., and L. Pritchett (2001). Estimating Wealth Effects without Expenditure Data – or Tears: With an Application to Educational Enrollments in States of India. *Demography*. Vol.38(1):115-132.

Forssas E, M. Gissle, M. Sihvonen, and W. Hemminki (1999). Maternal predictors of perinatal mortality: the role of birth weight. *International Journal of Epidemiology*, Vol. 28(3): 475-478.

Fortson J. G. (2008). The Gradient in Sub-Saharan Africa: Socioeconomic Status and HIV/AIDS. *Demography*. Vol. 45(2): 303–322.

Fowler M. G, and M.L. Newell (2002). Breast-feeding and HIV-1 transmission in resource-limited settings. *Journal of Acquired Immune Deficiency Syndrome*. Vol. 30(2): 230–239.

Fuchs, V. R. (1966). The Contribution of Health Services to the American Economy. *Milbank Memorial Fund Quarterly.* Vol.44(4): 65-102.

Fukuda-Parr, S. (1999). What Does Feminization of Poverty Mean? It Isn't Just Lack of Income. *Feminist Economics, Taylor and Francis Journal*. Vol.5(2): 99-103.

Gabriele B. D. (2005). Imputation Methods for Handling Item-Nonresponse in the Social Sciences: A Methodological Review. National Centre for Research Methods Review Paper NCRM/002. University of Southampton, UK.

Galbraith, J. K., (1996). The good society: The humane agenda. *Houghton Mifflin Co.* New York, USA. Pp.152.

Gertler, P., and P. Glewwe (1992). The Willingness to Pay for Education for Daughters in Contrast to Sons: Evidence from Rural Peru. *The World Bank Economic Review,* Vol.6(1), 171 - 188.

Gillespie S. R., S. Kadiyala and R. Greener. (2007). Is poverty or wealth driving HIV transmission? *AIDS.* Vol. 21 (7):5 – 16.

Glewwe, P. and H. O. Jacoby, (1994). Student Achievement and Schooling Choice in Low-Income Countries. *Journal of Human Resources,* Vol. 28(3), 843 - 864.

Gomes, M. (1984). Family Size and Educational Attainment in Kenya. *Population and Development Review,* Vol.10(4):647 - 660.

Goux D. and E., Maurin. (2005). The effect of overcrowded housing on children's performance at school, *Journal of Public Economics*, Elsevier. Vol.89(5-6):797-819

Green L.W., and J.M. Ottoson (1999). Community and population health. 8th ed. *McGraw-Hill*. Boston, USA. pp. 784.

Greene, W. (2003). Econometric Analysis, 5th Edition, *Prentice Hall*. New Jersey, USA. Pp.1026.

Grissom J., and L. A. Shepard (1989). Repeating and Dropping Out of School. In: Flunking Grades: Research and Policies on Retention, Shepard L. A., & M. L. Smith (Eds) pp. 34 -63. *Falmer Press*, London, UK.

Grossman M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, Vol. 80(2): 223-255.

Grossman M., and J. Joyce (1990). Unobservables, Pregnancy Resolutions, and Birth Weight Production Functions in New York City, *Journal of Political Economy.* Vol.98(5): 983-1007.

Guerin P., B. Fatuma, H. Elmi, and C. Corrigan (2007). Body Composition and Cardiorespiratory Fitness among Refugee Somali Women Living in New Zealand. *Journal of Immigrant and Minority Health.* Vol. 9(3): 191-196.

Halfdanarson, B., D. F. Heuermann, and J. Sudekum (2008). Human Capital Externalities and the Urban Wage Premium: Two Literatures and their Interrelations. *IZA Discussion Paper No. 3493*. Institute for the Study of Labor (IZA), Germany.

Hallett T. B., J. Stover, V. Mishra, P. D. Ghys, S. Gregson, and T. Boerma (2010). Estimates of HIV incidence from household-based prevalence surveys. *AIDS Official Journal of The International AIDS Society.* Vol. 24(1):147–152.

Hamill P.V., T.A. Drizd, C.L. Johnson, R.B. Reed, and A.F. Roche (1977). NCHS growth curves for children birth–18 years. Vital and Health Statistics Series 11, No 165. *National Center for Health Statistics*, U.S. Department of Health and Human Services, Hyattsville, MD. USA.

Handa, S. (1996). Maternal Education and Child Attainment in Jamaica: Testing the Bargaining Power Hypothesis, *Oxford Bulletin of Economics and Statistic,* Vol.58(1):119-137.

Hardin, J. W. (1996). sg61: Bivariate probit models. Stata Technical Bulletin 33: 15–20. Reprinted in *Stata Technical Bulletin Reprints*. *College Station, TX: Stata Press.* Vol. 6(33):152–158.

Hargreaves, J. R., and J. R. Glynn (2002). Educational attainment and HIV-1 infection in developing countries: a systematic review. Tropical Medicine and International Health. Vol 7(6):489–498.

Hart S.N., and M. R. Brassard (1987). A major threat to children's mental health: psychological maltreatment. *American Psychologist.* Vol.42(2):160-165.

Heckman J.J. (1979). Sample Selection Bias as a Specification Error, *Econometrica*, Vol. 47(1): 153–62.

Heckman, J. J. (1978). Dummy Endogenous Variables in a Simultaneous Equation System, *Econometrica,* Vol. 46(93): 1-59.

Hill A. (1992). Trends in childhood mortality in sub-Saharan mainland Africa, pp. 10-31, in E. van de Walle, P. Gilles, and M. Sala-Diakanda (eds.), Mortality and society in sub-Saharan Africa. *Clarendon Press*. Oxford, UK.

Hill K., and D. M. Upchurch (1995). Gender Differences in Child Health: Evidence from the Demographic and Health Surveys. *Population and Development Review*. Vol. 21(1): 127-151.

Hill K., G. Bicego, and M. Mahy (2001). Childhood Mortality in Kenya: An Examination of Trends and Determinants in the Late 1980s to Mid-1990s. *Johns Hopkins Population Center*. Working Paper *01-01*, Johns Hopkins School of Public Health, Baltimore, USA.

Hill, R. C., and M. S. Waters (1995). Estimation of a simultaneous Equations Model With an Ordinal Dependent Variable: The Extent of Teacher Bargaining and the State Legal Environment. *Journal of Economics and Finance*. Vol.19(5): 45-63.

Hobcraft J. (1993). Women's education, child welfare and child survival: a review of the evidence. *Health Transition Review*; Vol. 3(2): 159-173.

Hobcraft J.N., J.W. McDonald, and S.O. Rutstein (1984). Socioeconomic factors in infant and child mortality: a cross-national comparison. *Population Studies* Vol. 38(2):193-223

Hoffman. B. (2002). Womanhood and Circumcision: Three Maasai Women Have Their Say. *Berkeley Media LLC*. Berkeley, USA.

Horton S. (1988). Birth Order and Child Nutritional Status: Evidence from the Philippines. *Economic Development and Cultural Change*. Vol. 36(2): 341-354

Hosseinpoor, A. R., K. Mohammad, R. Majdzadeh, M. Naghavi, F. Abolhassani, A. Sousa, N. Speybroeck, H. R. Jamshidi, and J. Vega (2005). Socioeconomic inequality in infant mortality in Iran and across its provinces. *Bulletin of the World Health Organization*. Vol. 83(11): 837-844.

Hudson, C. P. (1996). AIDS in rural Africa: a paradigm for HIV-1 prevention. *International Journal of Sexually Transmitted Diseases and AIDS.* Vol. 7(4):236 - 43.

Hunt, R. (2010). Virology – Chapter Seven- human immunodeficiency virus and aids. Microbiology and immunology on–line, *University of Carolina: School of Medicine*. Retrieved Jan. 2010, from *http://pathmicro.med.sc.edu/lecture/hiv2000.htm*

Hussain T.M. (2004). HIV/AIDS in central Asia: The issue of a serious concern. *International Conference on AIDS* [Ref E10169]. Bangkok, Thailand.

Ibrahim, J. G., M. H. Chen, S. R. Lipsitz, and A. H. Herring (2005). Missing-Data Methods for Generalised Linear Models: A Comparative Review. *Journal of the American Statistical Association*. Vol. 100(469): 332-346.

Jacoby, H. 0. (1994). Borrowing Constraints and Progress through School: Evidence from Peru. *Review of Economics and Statistics.* Vol.76(1): 151 - 160.

Jonathan L. Heeney, Angus G. Dalgleish and Robin A. Weiss (2006) Origins of HIV and the Evolution of Resistance to AIDS. *Science.* Vol. 313(5786):462-466

Jaffe, P., D. Wolfe, and S.K. Wilson (1990). Children of Battered Women. Developmental Clinical Psychology and Psychiatry. Vol. 21. *SAGE publications, International and Professional Publisher,* Newbury Park, California, USA. Pp. 132.

Joint United Nations Programme on HIV/AIDS [UNAIDS] (2004). Report on the global AIDS epidemic. *Joint United Nations Programme on HIV/AIDS (UNAIDS),* Geneva, Switzerland.

Kabubo-Mariara, J., G. K. Ndenge, and D. K. Mwabu, (2009). Determinants of Children's Nutritional Status in Kenya: Evidence from Demographic and Health Surveys. *Journal of African Economies,* Vol. 18(3):363-387.

Kabubo-Mariara J. and D. K. Mwabu (2007). Determinants Of School Enrolment And Education Attainment: Empirical Evidence From Kenya. *South African Journal of Economics, Economic Society of South Africa*. Vol.75(3):572-593

Kaestner, R. (1997). Are Brothers Really Better: Sibling Sex Composition and Educational Achievement Revisited. *Journal of Human Resources*, Vol.32(2):250-284.

Kaiser H. F. (1970). A second generation little jiffy. *Psychometrika,* Vol. 35 (4): 401-416.

Kaiser H. F., and J. Rice (1974). Little jiffy mark IV. *Journal of Educational and Psychological Measurement,* Vol.34(1): 111-117.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement,* Vol. 20: 141-151.

Kanogo, T., (1987). Squatters and the Roots of Mau Mau. *East African Educational Publishers Ltd,* Westlands, Nairobi, Kenya. Pp. 224.

Keele, B.F., F. Van Heuverswyn, *et al*., (2006). Chimpanzee Reservoirs of Pandemic and Nonpandemic HIV-1, *Science,* Vol.313(5786):523-526.

Kenya Information guide. The Luo tribe. Retrieved Dec. 2010, from *http://www.kenya-information-guide.com/luo-tribe.html*

Kenya National Bureau of Statistics [KNBS] (2009). Kenya 2009 Population and Housing Census Highlights. *Kenya National Bureau of Statistics.* Nairobi, Kenya.

Kenyatta, J., (1979). Facing Mount Kenya. Vintage Books, A Division of Random House, New York, USA.

Kimalu, P.K., N. Nafula, D.K. Manda, A.S. Bedi, G. Mwabu, and M.S. Kimenyi (2001). Education indicators in Kenya. *KIPPRA Working Paper No. 4. Nairobi: Social Sector Division, Kenya Institute for Public Policy Research and Analysis.* Bishops Garden Towers, Nairobi, Kenya.

Kimalu, P.K., N. Nafula, D.K. Manda, A.S. Bedi, G. Mwabu, and M.S. Kimenyi (2001). A situational Analysis of Poverty in Kenya. *KIPPRA Working Paper No. 6. Nairobi: Kenya Institute for Public Policy Research and Analysis.* Bishops Garden Towers, Nairobi, Kenya

Knight J. B., and R. H. Sabot. (1990). Education, Productivity, and Inequality: The East African Natural Experiment. *World Bank*, Washington, DC. USA.

Knight, J. B., R. H. Sabot, and D. C. Hovey (1992). Is the Rate of Return on Primary Schooling Really 26 Per Cent? *Journal of African Economies, Vol.*1(3): 192–205.

Knodel, J., N. Havanon, and W. Sittirai, (1990). Family Size and the Education of Children in the Context of Rapid Fertility Decline. *Population and Development Review, Vol.*16(1):31 - 62.

Kowalewski, M. R. (1990). Religious Construction of the AIDS Crisis. Sociology of Religion. Vol. 51 (1): 91-96.

Kuczmarski R.J., C.L. Ogden, S.S. Guo, L.M. Grummer-Strawn, K.M. Flegal, Z. Mei, R. Wei, L. R. Curtin, A.F. Roche, and C. L. Johnson (2002). 2000 CDC growth charts for the United States: Methods and development. *National Center for Health Statistics Vital Health Stat.* Vol. 11(246): 1- 190.

Kulaga Z., M. Litwin, M. Tkaczyk, A. Rozdzynska, K. Barwicka, A. Grajda, A. Swiąder, B. Gurzkowska, E. Napieralska, and H. Pan (2010). The height-, weight-, and BMI-for-age of Polish school-aged children and adolescents relative to international and local growth references. *BioMed Central Public Health*, Vol. 10:109.

Kurian, G. T. (1992). Encyclopedia of the Third World, fourth edition, volume III, Facts on File: New York, N.Y. USA. Pp. 970-71 and 984-985.

Kutty N. K. (2000). The Production of Functionality by the Elderly: A Household Production Function Approach, *Applied Economics* Vol. 32(10): 1269-1280.

Kutty N. K. (2008). A Household Production Function Model of the Production of Child Health and Education - The Role of Housing-Related Inputs. Available at SSRN: *http://ssrn.com/abstract=1232602.*

Lagarde E., M. Schim, C. Enel, B. Holmgren, R. Dray-Spira, G. Pison, J. P. Piau, V. Delaunay, S. M'Boup, I. Ndoye, M. Coeuret-Pellicer, H. Whittle, and P. Aaby (2003). Mobility and the spread of human immunodeficiency virus into rural areas of West Africa. *International Journal Epidemiology*, Vol.32(5):744–752.

Lalonde M. (1974). A New Perspective on the Health of Canadians: A Working Document. *Ottawa, Government of Canada*, 1974. pp. 77.

Lau C., and A. S. Muula (2004). HIV/AIDS in Sub-Saharan Africa. *Croatian Medical Journal*, Vol. 45(4):402-414.

Law C. M., P Egger, O. Dada, H. Delgado, E. Kylberg, P. Lavin, G.H. Tang, H. von Hertzen, A.W. Shiell, and D.J.P. Barker (2001). Body size at birth and blood pressure among children in developing countries. *International Journal of Epidemiology*, Vol. 30(1): 52-57.

Leigh J.P. (1990). Schooling and Seat Belt Use. *Southern Economic Journal,* Vol. 57(1):195-207.

Little, R. J. A. and D. B. Rubin, D.B. (1990). The Analysis of Social Science Data with Missing Values. *Sociological Methods and Research,* Vol. 18(3): 292-326.

Loaiza, E. and C.B. Lloyd. (2008). Adolescents and Education in Africa. *Division of Policy and Practice, UNICEF*, Population Council.

Lucas, R.E. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics,* Vol.22(1):3-42.

Lugalla, J., M. Emmelin, A. Mutembei, M. Sima, G. Kwesigabo, J. Killewo, and L. Dahlgren (2004). Social, cultural and sexual behavioral determinants of observed decline in HIV infection trends: lessons from the Kagera Region, Tanzania. *Social Science and Medicine*, Vol.59(1): 185-198.

Luginnah, I., D. Elkins, E. Maticka-Tyndale, T. Landry, and M. Mathui (2005). Challenges of a pandemic: HIV/AIDS related problems affecting Kenyan widows. *Social Science & Medicine,* Vol.60 (6):1219-28.

Macassa G., G. Ghilagaber, E. Bernhardt, and B. Burstrom (2003. Trends in infant and child mortality in Mozambique during and after a period of conflict. *Journal of the institute of public health,* Vol.117(4):221-227.

Maddala, G. S. (1983). Limited-Dependent and Qualitative Variables in Econometrics. *Cambridge University Press.* Cambridge, UK. Pp. 401.

Madise N.J., and I. Diamond (1995). Determinants of infant mortality in Malawi: an analysis to control for death clustering within families. *Journal of Biosocial Science,* Vol. 27 (1): 95-106.

Maitra, P. (2003). Schooling and Educational Attainment: Evidence from Bangladesh. *Education Economics,* Vol.11(2):129-153.

Manda, D. K., G. Mwabu, and M. S. Kimenyi (2006). Human Capital Externalities and Private Returns to Education in Kenya. *Eastern Economic Journal,* Vol. 32(3): 493-513.

Massimiliano, B. and R. Leombruni. (2009). Local Human Capital Externalities and Wages at the Firm Level: The Case of Italian Manufacturing. IZA Discussion Papers 4613. *Institute for the Study of Labor (IZA).*

Mbugua, G.G., L.N. Muthami, C.W. Mutura, S.A. Oogo, P.G. Waiyaki, C.P. Lindan, and N. Hearst. (1995). Epidemiology of HIV infection among long distance truck drivers in Kenya. *East African Medical Journal,* Vol. 72(8):515-518.

McElroy P.D., F.O. Ter Kuile, A.W. Hightower, W.A. Hawlley, P.A. Phillips-Howard, A.J. Oloo, A.A. Lal, and B.L. Nahlen. (2001). All cause mortality among young children in western Kenya. *American Journal of Tropical Medicine and Hygiene,* Vol.64(1-2 Suppl): 18-27.

Miller, D. (2010). Children's Exposure to Violence - Community Violence, Domestic Violence - GENERAL EFFECTS. Retrieved 10th January 2011. *http://education.stateuniversity.com/pages/2531/Violence-Children-s-Exposure.htm*

Mincer, J. A. (1974). Schooling, Experience, and Earnings. NBER Books, *National Bureau of Economic Research, Inc., Columbia Univerity Press,* New York, USA.

Ministry of Education [MoE], Human Resources Development [HRD], World Bank [WB] and Ministry of Health [MOH] (Kenya) (2001). *AIDS in Kenya.* Nairobi Kenya.

Ministry of Education Science and Technology [MEST], (Kenya), (1996). National report on the development of Education 1995 to 1996 in Kenya, *International Conference on Education (ICE), Forty-Fifth Session*. Geneva.

Ministry of Education Science and Technology [MEST], (Kenya), (1998). Master plan on Education and Training, *Government Printer.* Nairobi, Kenya.

Ministry of Education Science and Technology [MEST], (Kenya), (2001). National report on the development of Education in Kenya, *International Conference on Education (ICE), Forty-Sixth Session.* Geneva,

Ministry of Education Science and Technology [MEST], (Kenya), (2004). Development of education in Kenya. Nairobi, Kenya.

Ministry of Education Science and Technology, [MEST], (Kenya), (1996). An overview on Human Development Through Education and Training in Kenya: Policies and Programmes Priorities, Nairobi Kenya.

Mishra V., S. Bignami-Van Assche, R. Greener, M. Vaessen, R. Hong, P. D. Ghys, J. T. Boerma, A. Van Assche, S. Khan, and S. Rutstein (2007). HIV infection does not disproportionately affect the poorer in sub-Saharan Africa. *AIDS,* Vol.21(7):17–28.

Mishra, V., F. Anold, F. Otieno, A. Cross, and R. Hong (2007). Education and Nutritional Status of Orphans and Children of HIV-Infected Parents in Kenya. AIDS Educational and Prevention, Vol.19(5):383-95.

Mishra, V., R. Hong, Y. Gu, A. Medley, and B. Robey (2009). Levels and Spread of HIV Seroprevalence and Associated Factors: Evidence from National Household Surveys. DHS Comparative Reports No. 22. *Calverton, Maryland, USA: Macro International Inc.*

Mondal Md. N.I., Md. K. Hossain, and Md. K. Ali (2009) Factors Influencing Infant and Child Mortality: A Case Study of Rajshahi District, Bangladesh. *Journal of Human Ecology*, Vol.26(1):31-39.

Montgomery, M.R., M. Gragnolati, K.A. Burke, and E. Paredes (2000). Measuring Living Standards With Proxy Variables. *Demography,* Vol.27(2):155–74.

Morduch, A. (2000). Sibling rivalry in Africa. *American Economic Review,* Vol.90(2):405–9.

Morris, C. N., and A G Ferguson (2006). Estimation of the sexual transmission of HIV in Kenya and Uganda on the trans-Africa highway: the continuing role for prevention in high risk groups. *Sexually Transmitted Infections*, Vol. 82(5):368–371

Moses, S., J.E. Bradley, N.J.D. Nagelkerke, A.R. Ronald, J.O. Ndinya-Achola, and F.A. Plummer (1990). Geographical patterns of male circumcision practices in Africa: Association with HIV seroprevalence. *Intenationla Journal of Epidemiology*, Vol.19(3):693-697.

Mosley, W., and L. Chen, (1984). An Analytical Framework for the Study of Child Survival in Developing Countries. *Population and Development Review.* Vol.10(supl): 25-45.

Mturi, A.J., and S.L. Curtis (1995). The determinants of infant and child mortality in Tanzania. *Health Policy Plan*, Vol.10(4):384-94.

Mukhopadhyay I., S. K. Bandyopadhyay and A. Chatterjee (2011). Prioritisation of the determinants of customer satisfaction: A simultaneous equation approach in ordinal endogenous set-up. *Total Quality Management & Business Excellence*. Vol.22(1):117-130.

Mungai, A.M. (2002). The quest for education in Post Colonial Kenya: My Personal experience. Paper presented at *the Annual Meeting of the Comparative and International Education Society*, Orlando, Florida. USA.

Munro, J.F. (1975). Colonial Rule and the Kamba. Social Change in the Kenya Highlands, 1889-1939. *Clarendon Press*. Oxford, UK. Pp.276.

Murray, C. J. L., and A. Lopez. (1996). The Global Burden of Disease: A comprehensive assessment of mortality from diseases, injuries and risk factors in 1990 and projected to 2020. *Harvard University Press*, Cambridge, Massachusetts

Mushkin, S. J. (1962). Health as an Investment. *Journal of Political Economy*, Vol. 70(5):129-38.

Mustafa, H. E., and C. Odimegwu (2008). Socioeconomic Determinants of Infant Mortality in Kenya: Analysis of Kenya DHS 2003. *Journal of Humanities & Social Sciences*, Vol 2(2):1-16.

Mutunga, C. J. (2007). Environmental Determinants of Child Mortality in Kenya. Working Papers UNU-WIDER Research Paper, *World Institute for Development Economic Research (UNU-WIDER).*

Mwabu, G. (2009). The Production of Child Health in Kenya: A Structural Model of Birth Weight. *Journal of African Economies.* Vol.18(2):212–260.

Nafziger, E. W. (2006). Economic Development, Fourth Edition. *Cambrige University Press*. Cambrige, UK. Pp. 868.

National AIDS/STD Control Programme [NASCOP] (2005). AIDS in Kenya: background, projection, impact, interventions and policy. *Nairobi: AIDS Control Unit, Ministry of Health, National Aids Control Council, Government printers*.

Nawata, K. (1994). Estimation of Sample-Selection Biases Models by the Maximum Likelihood Estimator and Heckman's Two-step Estimator, *Economic Letters*, Vol. 45(1):33-40

Nelson, F. D., and L. Olsen, (1978). Specification and Estimation of a Simultaneous Equation Model with Limited Dependent Variables, *International Economic Review,* Vol. 19(3): 695-709.

Newell, M.L., H. Brahmbhatt, and P.D. Ghys (2004). Child mortality and HIV infection in Africa: a review. *AIDS.* Vol.18 (2):27–34.

Nii-Amoo Dodoo, F., E.M. Zulu, A.C. Ezeh (2007). Urban–rural differences in the socioeconomic deprivation – sexual behavior link in Kenya. *Social Science & Medicine,* Vol. 64 (5):1019–1031

Nugent, J. B., (1985). The Old-Age Security Motive for Fertility. *Population and Development Review,* Vol.11(1):75-97.

Nugent, J. B., and T. R. Gillaspy (1983). Old Age Pensions and Fertility in Rural Areas of Less Developed Countries: Some Evidence from Mexico. *Economic Development and Cultural Change,* Vol.31(4):809-829.

Nzyuko, S., P. Lurie, W. McFarland, W. Leyden, D. Nyamwaya, and J.S. Mandel (1997). Adolescent sexual behavior along the Trans-Africa Highway in Kenya. *AIDS.* Vol. 11 (1):21–26

Nzyuko, S., R.W. Morgan, and D. Nyamwaya (1991). AIDS and condoms: some notes on the "invisible" disease. *Africa Urban Quarterly,* Vol. 6(1&2):59-73.

Obermeyer, C.M., (2003). The health consequences of female circumcision: Science, advocacy, and standards of evidence. *Medical Anthropology Quarterly,* Vol.17(3):394-412.

Obonyo, B., F. Otieno, and R.O. Muga (2005). Effect of Infant and Child Mortality on Fertility in Kenya, Working Paper No. 2, Nairobi, Kenya: National Coordinating Agency for Population and Development.

Oddy, W.H, G.E. Kendall, J. Li, P. Jacoby, M. Robinson, N.H. de Klerk, S.R. Silburn, S.R. Zubrick, L.I. Landau, and F.J. Stanley (2010). The long-term effects of breastfeeding on child and adolescent mental health: a pregnancy cohort study followed for 14 years. *The Journal of Pediatrics*, Vol.156(4):568-574.

Olson, J.B. (1972). Secondary Schools and Elites in Kenya: A comparative Study of Students in 1961 and 1968. *Comparative Education Review*, Vol.16(1):44-53.

Omariba, D., R. Beaujot, and F. Rajulton (2007). Determinants of infant and child mortality in Kenya: an analysis controlling for frailty effects. *Population Research and Policy Review*, Vol.26(3):299-321.

Onis, M, M. Blossner, E. Borghi, E.A. Frongillo and R. Morris (2004). Estimates of Global Prevalence of Childhood Underweight in 1990 and 2015. *Journal of the American Medical Association*, Vol.291(21):2600-2606.

Onis, M., and M. Blossner (2003). The World Health Organization Global Database on Child Growth and Malnutrition: methodology and applications. *International Journal of Epidemiology*, Vol.32(4):518-526.

Oreopoulos, P., M. Page, and A.H. Stevens (2003). Does Human Capital Transfer from Parent to Child? The Intergenerational Effects of Compulsory Schooling. NBER Working Paper 10164. *National Bureau of Economic Research, Inc.*

Page, M. (2006). Fathers' Education and Children's Human Capital: Evidence from the World War II G.I. Bill. Working paper 06-33, *University of California*, Davis.

Parish, W.L., and R.J. Willis (1993). Daughters, Education, and Family Budgets: Taiwan Experiences, *Journal of Human Resources*, Vol.28(4):863-898.

Parkhurst, J.O. (2010). Understanding the correlations between wealth, poverty and human immunodeficiency virus infection in African countries. *Bulletin of the World Health Organization*, Vol.88(7):481-560

Pine, C.M. (1997). Introduction, principles and practice of public health. In: Pine CM, editor. Community oral health. *Reed Educational and Professional Publishing*, Oxford. pp. 1-10.

Posel, D.R. (2001). Who are the heads of household, what do they do, and is the concept of headship useful? An analysis of headship in South Africa. *Development Southern Africa,* Vol.18(5):651-670.

Psacharopoulos, G. (1985). Returns to Education: A Further International Update and Implications. *Journal of Human Resources,* Vol.20:583–604.

Psacharopoulos, G., (1997). Child Labour versus Educational Attainment: Some Evidence from Latin America. *Journal of Population Economics,* Vol.10(4):377-386.

Psacharopoulos, G., and A. M. Arriagada (1989). The Determinants of Early Age Human Capital Formation: Evidence from Brazil. *Economic Development and Cultural Change, Vol.*37(4):683-708.

Psacharopoulos, G., and M. Woodhall (1985). Education for Development: An Analysis of Investment Choice. *Oxford University Press,* New York. Pp., 337.

Psacharopoulos,G. (1994). Returns to Investment in Education: A Global Update. *World Development, Elsevier,* Vol.22(9):1325–1343.

Rahman M. (2009). Factors Affecting on Child Survival in Bangladesh: Cox Proportional Hazards Model Analysis. *The Internet Journal of Tropical Medicine.* Vol. 6(1).

Rahman, M, S.K. Roy, M. Ali, A.K. Mitra, A.N. Alam, and M.S. Akbar (1993) Maternal Nutritional Status as a Determinant of Child Health. *Journal of Tropical Pediatrics,* Vol.39(2):86-88.

Ray, D. (1998). Development Economics, *Princeton University Press*. Princeton, New Jersey UK. Pp 848.

Rivers, D., and Q.H. Vuong (1988). Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models. *Journal of Econometrics,* Vol. 39: 347 — 366.

Roberts, D.F, and D.R. Bainbridge (1963). Nilotic physique. *American Journal of Physical Anthropology*. Vol. 21(3):341-370.

Romer, P. M, (1990). Endogenous Technological Change. *Journal of Political Economy, University of Chicago Press*, Vol. 98(5):S71-S102.

Romer, P. M. (1986). Increasing Returns and Long-run Growth. *Journal of Political Economy, University of Chicago Press,* Vol.94(5):1002-1037.

Rosenzweig, M.R., and P.T. Schultz (1983). Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight. *Journal of Political Economy,* Vol.91(5):723-746.

Rowley, E.A., P.B. Spiegel, Z. Tunze, G. Mbaruku, M. Schilperoord, and P. Njogu (2008). Differences in HIV-related behaviors at Lugufu refugee camp and surrounding host villages, Tanzania. *Conflict and Health.* Vol.2(13).

Royston, P. (2004). Multiple imputation of missing values. *Stata Journal*, Vol.4(3): 227-241.

Royston, P. (2005). Multiple imputation of missing values: update of ice. *Stata Journal*, Vol.5(4):527-536.

Royston, P. (2005). Multiple imputation of missing values: update. *Stata Journal,* Vol.5(2):188-201.

Rubin, D. B (1987). Multiple imputation for non-response in surveys. *John Wiley & Sons*, New York. Pp. 288.

Rubin, D.B. (1976). Inference and missing data (with discussion). *Biometrika,* Vol.63(3): 581–592.

Ruger, J.P., and H-J Kim (2006). Global health inequalities: an international comparison. *Journal of Epidemiol Community Health,* Vol.60(11):928-936.

Sanghvi, T., and J. Murray (1997). Improving child health through nutrition: The nutrition Minimum Package. *Arlington, Va.: Basic Support for Institutionalizing Child Survival (BASICS) Project, for the U.S.* Agency for International Development.

Sateren, W.B., G. Foglia, P.O. Renzullo, L.E., M. Wasunna, C.T. Bautista, and D.L. Birx (2006). Epidemiology of HIV-1 Infection in Agricultural Plantation Residents in Kericho, Kenya: Preparation for Vaccine Feasibility Studies. *Journal of Acquired Immune Deficiency Syndromes,* Vol. 43(1):102 -106.

Schafer, J.L., and J.W. Graham (2002). Missing Data: Our View of the State of the Art, *Psychological Methods.* Vol.7(2):147-177.

Schotz, J.T, and M. Lubell (1998). Trust and Taxpaing: Testing the heuristic approach to collection action. *American Journal of Political Science,* Vol.42(2):398-417.

Schultz, T. W. (1961). Investment in Human Capital. *American Economic Review*, Vol. 51(1):1-17.

Scott, B.E., H.A. Weiss, and J.I. Viljoen (2005).The acceptability of male circumcision as an HIV intervention among a rural Zulu population, Kwazulu-Natal, South Africa. *AIDS Care*, Vol. 17(3):304-13.

Shavit, Y., and J. L. Pierce (1991). Sibship Size and Educational Attainment in Nuclear and Extended Families: Arabs and Jews in Israel. *American Sociological Review, Vol.*56(3):321-330.

Shea, J. (1997). Instrument Relevance in Multivariate Linear Models: A Simple Measure, *Review of Economics and Statistics.* Vol.49(2):348–52.

Shelton, J.D., and B. Johnston (2001). Condom gap in Africa: evidence from donor agencies and key informants. *British Medical Journal,* Vol. 323(7305):139.

Shelton, J.D., M.M. Cassell, and J. Adetunji (2005). Is poverty or wealth at the root of HIV? *Lancet,* Vol.366(9491):1057-8.

Sherif, N. (1991). On the computation of a matrix inverse square root. *Computing,* Vol. 46(4):295-305

Shlens, J. (2005). A Tutorial on Principal Components Analysis. *Institute for Nonlinear Science, University of California*, San Diego La Jolla, CA.

Sigmund, T., and J.E. Carlson (1969). Brief Report: Bartlett's Test of Sphericity and Chance Findings in Factor Analysis. *Multivariate Behavioral Research*, Vol.4(3):375–377.

Smith, P. (2006). Kenya First Lady: Condom is causing the spread of AIDS in this country, Website: *LifeSiteNews.com /ldn/2006/may/06052307.html.*

Snedecor, G.W., and W.G. Cochran (1989). Statistical Methods, 8th Ed. *Iowa State University Press*. Pp. 503

Spence, M. (1973). Job marketing signalling. *Quarterly Journal of Economics,* Vol.87(3):388-374.

STATA/IC 11.1 for Windows, (2010). *Stata Corporation*, College Station, Texas, USA.

Staten, L.K., D.L. Dufour, J.C. Reina, and G.B. Spurr (1999). Household headship and nutritional status: female-headed versus male-dual-headed households. *American Journal of Human Biology*, Vol.10(6):699-709.

Staudt, A.K. (1985). Agricultural policy implementation: A case study from western Kenya. *Kumarian Press*, West Hartford, Connecticut, USA. Pp. 68.

Stiglitz, J. E. (1975). The theory of 'screening', education and distribution of income. *American Economic Reviews,* Vol.65(354):283-300.

Stock, J.H., H.W. Jonathan and M. Yogo (2002). A Survey of Weak Instruments and Weak Identification in Generalized Methods of Moments. *Journal of Business and Economic Statistics*, Vol.20(4):518–29.

Suzanna, V., J. Carlin, K. Hesketh, and T. Cole (2004). Standardizing anthropometric measures in children and adolescents with new functions for egen. *The Stata Journal*, Vol.4(1):50–55.

Szabo, R., and R.V. Short, (2000). How Does Male Circumcision Protect Against HIV Infection? *British Medical Journal*, Vol.320(7249):1592-1594.

Taha, T.E., J.K. Canner, J.D. Chiphangwi, *et al.*, (1996). Reported condom use is not associated with incidence of sexually transmitted diseases in Malawi. *AIDS.* Vol.10(2):207-212.

Taha, T.E., G.A. Dallabetta, D.R. Hoover, J.D. Chiphangwi, L.A. Mtimavalye, G.N. Liomba, N.I. Kumwenda, and P.G. Miotti (1998). Trends of HIV-1 and sexually transmitted diseases among pregnant and postpartum women in urban Malawi. *AIDS*, Vol.12(2):197-203.

Tansel, A. (1997). Determinants of School Attainment of Boys and Girls in Turkey, *Working Paper No. 789, Economic Growth Center, Yale University.* New Haven CT, USA

Tanzania Commission for AIDS [TACAIDS] (2005). National Bureau of Statistics (NBS), and ORC Macro. *Tanzania HIV/AIDS Indicator Survey 2003–04*. Calverton, Maryland USA: TACAIDS, NBS, and ORC Macro.

The Joint United Nations Programme on HIV/AIDS [UNAIDS] (2006). Chapter 3: Progress in Countries. In: Report on the Global AIDS Epidemic: pages 52-78. *UNAIDS*, Geneva, Switzerland.

The Joint United Nations Programme on HIV/AIDS [UNAIDS], the United Nations Children's Fund [UNICEF], and the United States Agency for International Development [USAID] (2004). Children on the brink 2004: a joint report of the new orphan estimates and a framework for action. *United Nations Children's Fund*. New York, USA.

The Well Project. HIV: *The Basics*. Retrieved Jan. 2010, from *http://www.thewellproject.org/en_US/HIV_The_Basics/index.jsp*

The World Fact book, Kenya. *Central Intelligence Angency (CIA)*. Retrieved 15 Nov. 2010, from *https://www.cia.gov/library/publications/the-world-factbook/geos/ke.html*

UNAIDS, (2000). Report on the Global HIV/AIDS Epidemic: pages 28-29. *UNAIDS*, Geneva, Switzerland.

UNAIDS (2009). AIDS epidemic update. *UNAIDS and World Health Organization (WHO)*. Geneva, Switzerland.

UNICEF (2003). Africa's orphaned generations. *United Nations Children's Fund,* New York, USA.

United Nations [UN] (1996). International year for the eradication of poverty: The geography of poverty. *United Nations Department of Public Information.*

United Nations Human Settlement Programme [UN–HABITAT] (2002). Management of HIV/AIDS Pandemic at the local level. *The United Nations Settlements Programme, UN-HABITAT,* Vol.8(2):19-20.

Van de Poel, E., A.R. Hosseinpoor, N. Speybroeck, T. Van Ourti, J. Vega (2008). Socioeconomic inequality in malnutrition in developing countries. *Bulletin World Health Organ,* Vol.86(4):282-291.

Van den, J.B, D. Willie, and N. Younger (2009). The World Health Organization child growth standards: expected implications for clinical and epidemiological research. *European Journal of Pediatrics,* Vol.168(2):247–251.

Van de Ven, W. P. M. M., and B. M. S. Van Pragg. (1981). The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of Econometrics,* Elsevier. Vol.17(2): 229–252.

Venniker, R. (2001). Social Returns of education, A survey of recent literature on Human Capital Externalities. CPB (Netherlands Bureau for Economic Policy Analysis) Report 00/1.

Wa-Githumo, M., (1981). Land and Nationalism. *Washington University Press of America,* Washington DC, USA. Pp.477.

Watson-Jones, D., K. Baisley, H. A. Weiss, C. Tanton, J. Changalucha, D. Everett, T. Chirwa, D. Ross, T. Clayton, and R. Hayes. (2009). Risk factors for HIV incidence in women participating in an HSV suppressive treatment trial in Tanzania. *AIDS,* Vol.23(3):415-22.

World Health Organization [WHO] (1946). Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York. USA.

WHO (1995). Physical Status: The Use and Interpretation of Anthropometry. Report of a WHO Expert Committee. Technical Report Series No. 854. *WHO*, Geneva, Switzerland.

WHO (2003). The World Health Report: Shaping the future. *WHO*, Geneva, Switzerland. Pp.193

WHO (2006). WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. *WHO*, Geneva, Switzerland. Pp. 312.

WHO (2007) WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards: Head circumference-for-age, arm circumference-for-age, triceps skinfold-for-age and subscapular skinfold-for-age: Methods and development. *WHO*, Geneva, Switzerland. Pp. 217.

WHO (2009). Investing in Maternal, Newborn and Child Health - The Case for Asia and the Pacific, World Health Organization (providing hosting arrangements for the PMNCH). *WHO*, Geneva, Switzerland. Pp. 10.

World Health Organization [WHO], United Nations Children's Fund [UNCEF] & Joint United Nations Programme on HIV/AIDS [UNAIDS] (2009). Towards universal access: scaling up priority HIV/AIDS interventions in the health sector. *WHO*, Geneva, Switzerland.

Wilson, K. (2001). The Determinants of Educational Attainment: Modeling and Estimating the Human Capital Model and Education Production Functions. *Southern Economic Journal*, Vol.67(3):518-551.

Woldemicael, G. (2001) Diarrhoeal morbidity among young children in Eritrea: Environmental and socio-economic determinants, *Journal of Health Population and Nutrition*, Vol.19(2):83-90.

Wolfe, N. D., W. M. Switzer, *et al.* (2004). Naturally acquired simian retrovirus infections in Central African Hunters' *The Lancet*, Vol.363(9413):932-937.

Wooldridge, J.M. (2006). Introductory Econometrics A Modern Approach, Ed 4. *Thompson South-Western*, USA. Pp.900.

World Bank [WB] (2006d). Repositioning Nutrition as Central to Development: A Strategy for Large-Scale Action. *World Bank.* Washington, DC, USA.

Wright, C.M., I.W. Booth, J.M.H. Buckler, N. Cameron, T.J. Cole, M.J.R. Healy, J.A. Hulse, M.A. Preece, J.J. Reilly, and A.F. Williams (2002). Growth reference charts for use in the United Kingdom. *Archives of Disease in Childhood,* Vol.86(1):11-14.

WUNRN. (2006). *KENYA: Cultural traditions fuel the spread of HIV/AIDS*. Retrieved from, *http://www.wunrn.com/news/2006/03_26_06/032706_kenya_ethnic.htm*

Yamano, T. & T S. Jayne (2005). Working-Age Adult Mortality and Primary School Attendance in Rural Kenya. *Economic Development and Cultural Change, University of Chicago Press.* Vol.53(3):619-53.

Zhou, H., R. Ohtsuka, Y. He, L. Yuan, T. Yamauchi, and A.C. Sleigh (2005). Impact of parasitic infections and dietary intake on child growth in the schistosomiasis-endemic Dongting Lake region, China. *American Journal of Tropical Medicine and Hygiene,* Vol.72(5):534–9.

Zozie, H. L. (1998). Contributing factors to high rate of HIV transmission among the communities of Lilongwe District in Malawi. *International Conference on AIDS,* Vol.12(1018).

Zuma, K., E. Gouws, B. Williams, and M. Lurie (2003). Risk factors for HIV infection among women in Carletonville, South Africa: migration, demography and sexually transmitted diseases. *International Journal of STD & AIDS,* Vol.14(12):814–817.