Logistic and Augmented Modeling of Poverty Profiles and Forecasting of the Food Crops Balance Sheet (Case of Lake Victoria Basin, Kenya)

Anthony Ngunyi

A thesis submitted in fulfillment for the Degree of Doctor of Philosophy in Applied Statistics in the Jomo Kenyatta University of Agriculture and Technology

2015

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University.

Signature: .... Date: ....

Antony Ngunyi

This thesis has been submitted for examination with our approval as University Supervisors.

Signature: .... Date: ....

Prof. Peter Nyamuhanga Mwita, PhD

JKUAT, KENYA

Signature: .... Date: ....

Prof. Romanus Otieno ODHIAMBO, Ph.D

JKUAT, KENYA

Signature: .... Date: 2nd Nov. 2015

Prof. Verdiana Grace Masanja, Ph.D

University of Rwanda, RWANDA

Affiliated to **University of Dar es Salaam**

DEDICATION

To my wife Nancy, my three children Collins, Joydrin and Abigael and my Mum Lydia.

ACKNOWLEDGEMENT

I would like to thank the almighty God for seeing me through this period of my entire study. I give my sincere gratitude to the Dedan Kimathi University of Technology for providing me with a scholarship to pursue this course at JKUAT.

I would like to thank more sincerely Prof. Peter N. Mwita, Prof. Romanus O. Odhiambo and Prof. Verdiana G. Masanja for tirelessly supervising the whole research work, their guidance and support is highly appreciated. I gratefully acknowledge the support I got from Dr. Waititu from JKUAT for his invaluable support and his various contributions to the success of this work.

The authors acknowledge the Inter-University Council for East Africa, under the Lake Victoria Research Initiative (VicRes), for providing financial support.

To my colleagues at work, Mr. Omari my chairman, Mr. Andrew Kinyita, Mr. Kamau Riro and others for their encouragement which made me always dream of the day I will have to achieve this important part of my life.

To my students in the Statistics and Actuarial Science department, you always gave me confidence and motivation to finish this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF SYMBOLS

$\mathbf{f}(\mathbf{y})$	Density function
$\mathbf{X}, \mathbf{Y}$	Random Variable
$\mathbf{x_i}, \mathbf{y_i}$	Realized values x, y lowercase (data)
$X = (\mathbf{x_1}, ..., \mathbf{x_n})$	Vector (boldface signifies vectors)
$\mathbf{H} = ((\mathbf{x_{ij}})) = \|\|\mathbf{x_{ij}}\|\|$	Matrix (uppercase signifies matrices)
$\mathbf{I}(.)$	Indicator function
$\mathbf{K}(\mathbf{u})$	Kernel function
$\mathbf{h}$	Smoothing parameter (“bandwidth”).
$\mathbf{fi} = \mathbf{fi_1}, \mathbf{fi_2}, \ldots \mathbf{fi_p}$	Parameter
$\mathbf{ß} = \mathbf{P}(\mathbf{Y} = \mathbf{1})$	Probability function
$\hat{\mathbf{fi}}$	Estimate
$\mathbf{L}(\mathbf{fi})$	Likelihood function
$\mathbf{I}(\mathbf{fi})$	Information Matrix
$\frac{\partial}{\partial \mathbf{x}}$	derivative
$\sum$	Variance-Covariance Matrix
$\mathbb{R}^{\mathbf{p}}$	Euclidean space p -dimension
$\sum_{\mathbf{i=1}}^{\mathbf{n}}$	Summation
$\mathbf{o}(\mathbf{n})$	Little “o”
$\mathbf{O}(\mathbf{n})$	Big “O”

ABBREVIATIONS

CBN	Cost of Basic Needs
FAO	Food Agriculture Organization
FBS	Food Balance Sheet
FEI	Food Energy Intake
FEWSNET	Famine Early Warning System Network
FGT	Foster, Greer and Thorbecke
FIVIMS	Food Insecurity and Vulnerability Information and Mapping Systems
GIEWS	Global Information and Early Warning System
HFI	Household Food Insecurity
IFPRI	International Forestly Resources and Institutes
KIHBS	Kenya Intergrated Housing Based Survey
KNBS	Kenya National Bureau of Statistics
MLE	Maximum Likelihood Estimator
NDP	National Development Plans
NPEP	National Poverty Eradication Plans
OLS	Ordinary Least Squares
PPAP	Participatory Poverty Alleviation Programs
PPP	Purschasing Parity Power
PPS	probability proportional to size
PRSP	Poverty Reduction Strategic Papers
SSR	Self Sufficiency Ratio
WFP	World Food Programme
WHO	World Health Organization
WMS	Welfare Monitoring Survey

ABSTRACT

The problem of poverty is one of the core issues concerning developing countries like Kenya. The formulation of an adequate programme to combat poverty is of importance for any meaningful development plan. The key features relevant are the construction of an appropriate poverty index and proper estimation of the measure. The different dimensions of poverty add to the problem of choosing the appropriate poverty measure and indicators. What is the appropriate measure to estimate the incidence of poverty? In other words, what criteria should be used to define and measure poverty? What is missing from previous studies is an analysis of different poverty measures. The study sought to propose a model that takes care of the multi-faceted nature of poverty and also look into the trends of food security in the Lake Victoria basin in three ways: Firstly, we come up with the poverty line of the region using the consumption data, secondly we look at the two models and estimate the best model for investigating poverty including a wide range of independent variables to reflect the contribution of each to a household being poor and lastly forecasting food insecurity using the food crops balance sheet. The assessment involved analysis using the augmented regression model and the stepwise model analysis for variable reduction, construction of logit models for different poverty proxies and application of the models in classification of households by poverty status. Further, assessment of poverty was made using assets, a multi-dimensional approach. Further analysis was done on the food balance sheet in order to obtain projections on food production and consumption patterns in the region. In the results we precisely state the asymptotic properties of maximum likelihood estimators for logistic regression models and additionally we show that the maximum likelihood estimators converge, under conditions of fixed number of predictor variables, to the real value of the parameters as the number of observations tends to infinity. The results also indicated that the parameters estimates are normal in distribution by plotting the quantile plots and undertaking the Kolmogorov-Smirnov and the Shapiro-Wilks test for normality, and conclude that parameters came from a normal distribution.The thesis comes up with some theoretical as well as empirical contributions taking into consideration various aspects of poverty measurement in the context of Lake Victoria basin, Kenya. A significant development for research has been the improvement in constructing a coherent framework for measuring poverty in multidimensional environment. This framework provides a new insight into particular elements of poverty that is useful and relevant to poverty interventions. The projections in this work are not statements of what will happen, but of what might happen, given the assumptions and methods used. These projections provide a policy-neutral starting point that can be used to analyze national and counties food requirement and policy initiatives.

CHAPTER ONE

INTRODUCTION

This chapter presents a detailed overview of the welfare and poverty concepts used in the thesis and outlines the measurement approaches adopted. Section 1.1 provides a description of the definition and construction of the welfare measure used to estimate poverty. Section 1.3 explains how different methods of constructing poverty lines, and details how the poverty lines used in this thesis were computed. Section 1.4 describes the food balance sheets as measures of food security in different regions and how they can be used to measure the state of welfare. Finally, section 1.5 -1.11 details the problem statement, justification of the problem and the scope of the study.

1.1 Background Information

Poverty is on the top of the agenda of policymakers and policy analysts around the world as it is both a cause and result of economic and social development. In fact, as argued in the Human Development Report (2000, 73) "eradication of poverty is more than a major development challenge - it is a human rights challenge". There are almost 1.5 billion people living in poverty. Perhaps more importantly, they are likely to remain in poverty for long periods of time. In other words, poverty is a persistent condition for some.

According to the WB (2010), "poverty is pronounced deprivation in well-being." This of course begs the questions of what is meant by well-being and of what is the reference point against which to measure deprivation.

One approach is to think of well-being as the command over commodities in general, so people are better off if they have a greater command over resources. The main focus is on whether households or individuals have enough resources to meet their needs. Typically, poverty is then measured by comparing individuals' income or consumption with some defined threshold below which they are considered to be poor. This is the most conventional view-poverty is seen largely in monetary terms-and is the starting point for most analysis of poverty.

A second approach to well-being (and hence poverty) is to ask whether people are able to obtain a specific type of consumption: Do they have enough food? Or shelter? Or health care? Or education? In this view the analyst goes beyond the more traditional monetary measures of poverty: Nutritional poverty might be measured by examining whether children are stunted or wasted; and educational poverty might be measured by asking whether people are literate or how much formal schooling they have received.

Perhaps the broadest approach to well-being is the one articulated by Sen (1986), who argues that well-being comes from a capability to function in society. Thus, poverty arises when people lack key capabilities, and so have inadequate income or education, or poor health, or insecurity, or low self-confidence, or a sense of powerlessness, or the absence of rights such as freedom of speech. Viewed in this way, poverty is a multidimensional phenomenon and less amenable to simple solutions. For instance, while higher average incomes will certainly help reduce poverty, these may need to be accompanied by measures to empower the poor, or insure them against risks, or to address specific weaknesses such as inadequate availability of schools or a corrupt health service.

Poverty is related to, but distinct from, inequality and vulnerability. Inequality focuses on the distribution of attributes, such as income or consumption, across the whole population. In the context of poverty analysis, inequality requires examination if one believes that the welfare of individuals depends on their economic position relative to others in society. Vulnerability is defined as the risk of falling into poverty in the future, even if the person is not necessarily poor now; it is often associated with the effects of "shocks" such as a drought, a drop in farm prices, or a financial crisis. Vulnerability is a key dimension of well-being since it affects individuals' behavior in terms of investment, production patterns, and coping strategies, and in terms of the perceptions of their own situations.

It takes time, energy, and money to measure poverty, since it can only be done properly by gathering survey data directly from households. Why, then, do we need to go to the trouble of measuring poverty? At least two good reasons come to mind.

Keeping Poor People on the Agenda

Perhaps the strongest justification is that provided by Ravallion (1998), who argues, "A credible measure of poverty can be a powerful instrument for focusing the attention of policy makers on the living conditions of the poor." Put another way, it is easy to ignore the poor if they are statistically invisible. The measurement of poverty is necessary if it is to appear on the political and economic agenda.

Targeting Domestic and Worldwide Interventions

A second reason for measuring poverty is to target interventions. Clearly, one cannot help poor people without knowing who they are. This is the purpose of a poverty profile, which sets out the major facts on poverty (and typically, inequality) and then examines the pattern of poverty to see how it varies by geography (for example, by region, urban/rural, mountain/plain), by community characteristics (for example, in communities with and without a school), and by household characteristics (for example, by education of household head, by size of household). A well-presented poverty profile is invaluable, even though it typically uses rather basic techniques such as tables and graphs, for a straight forward example, see Prescott and Pradhan (1997).

Poverty has been a major concern of many governments world over and many poverty reduction programs have been developed over time and across regions. Despite these efforts, poverty continues to be a key impediment to both human and economic prosperity Organisation for Economic Cooperation and Development(OECD, 2000). As put in various policy documents, for instance, the Sessional Paper No. 10 of 1965 on African Socialism and its Application to Planning in Kenya, the Government of Kenya directed its efforts to fighting poverty, disease and ignorance as part of its development objectives.

Consequent to this, various policies National Development Plans (NDP), Participatory Poverty Alleviation Programs (PPAP), National Poverty Eradication Plans (NPEP) and Poverty Reduction Strategic Papers (PRSP) have spelt out strategies to fight poverty. These policies notwithstanding, poverty levels have continued to increase. For instance in 1971, the number of Kenyans regarded as poor was 3.7 million, increasing to 11.5 million in 1994 and further to 13.3 million in 1997. According to the Welfare Monitoring Survey (WMS) of 1994, the incidence of poverty in Kenya was 47% in the rural areas and 29% in the urban areas. The absolute poverty line was Kshs. 980 per capita per month for rural areas and Kshs.1490 per capita for the urban areas. This has since increased to Kshs. 1239 and 2648 for the urban and rural areas respectively. Mwabu *et al.* (2002) estimated that the number of poor Kenyans had shot up to 15 million (about 56% of the total populations) in the year 2000.

The government has also established causes, constraints and the processes that engender and entrench poverty but despite these positive developments, poverty alleviation has remained elusive particularly from the 1980s. Poor economic performance has led to increased absolute poverty, i.e., people without adequate food and nutrition, inadequate access to basic services, education, health facilities, safe water and decent housing. This has been blamed on poor policy formulation, initiation, planning and implementation. WB (2007) categorizes poverty depending on the approach used in measuring it. For example, the income based definition of poverty seeks to specify a level of income per capita in a household below which the basic needs of the family cannot be met. However, it does not acknowledge variations in costs of similar goods for different consumers. The vital importance of non-market household production and non-monetarised exchanges in poor families is not counted.

The basic needs approach involves specifying a set of minimal conditions of life, usually the quality of the dwelling place, degree of crowding, nutrition adequacy and water supply. The proportion of the population lacking these conditions is used to estimate the degree of poverty. The advantage of this approach is that different conditions appropriate to different settings can be specified. However this reduces comparability of estimates in different situations. Similarly, it does not take into account the willingness of people to accept various tradeoffs deliberately, for example, a lower quality of dwelling for reduced transportation time and expense at work. Despite the many facets of defining poverty, WB (2000) admits that we have misconceptions about the poor, why they are poor and what is needed to help them out of this vicious cycle. Regardless of the many definitions of poverty and its multidimensional perspective, we can conclude that overall poverty takes many forms including lack of income and productive resources to ensure sustainable livelihood, hunger and malnutrition, ill health, limited access to education and other basic services, increased morbidity and mortality rates, homelessness and inadequate housing, unsafe environments, social discrimination and exclusion. It is also characterized by lack of participation in decision-making in civil, political, social and cultural life.

According to the participatory poverty assessment study in Tanzania WB (1997), wealth is associated with the ability to meet basic needs, particularly food. In that study, poverty was associated with skipping meals, cutting meals to one or two per day, involuntary changing diets, sending children to eat at neighbour's homes, and children performing poorly in schools as hunger makes them skip classes and affects their attention in class.

Probably the most important operational use of the poverty profile is to support efforts to target development resources toward poorer areas. However, which regions should command priority in targeting? This question has been answered at a highly aggregate level by most survey data (Kenya is a low-income country, (WB, 2010)) whose 38.8 million population had, on average, an annual income of 1,560 US$ Purschasing Parity Power (PPP) (Appendix 1). The prevailing macro-economic conditions between 2003 and 2008 have helped to improve the welfare of Kenyans. The economy grew at a sustained rate between 5% and 7% and only in 2008, due to the effects of the financial and economic crisis, did the economic growth rate drop to 1.7%. The national absolute poverty declined from 52.3% in 1997 to 46.1% in 2005/06 (KNBS, 2007b,c). In rural areas, overall poverty declined from 52.9% to 49.1%, while, in urban areas, poverty declined from 49.2% in 1997 to 38.8% over the same period. Despite the impressive gains in economic growth prior to the 2008 crisis poverty remains a major challenge. The Kenyan poverty profile reveals strong regional disparities in the distribution of poverty. According to the 2005/2006 survey, the lowest incidence of rural poverty was in Central province (30.3%), followed by Nyanza (47.9%), Rift Valley (49.7%), Eastern (51.1%), Western (53.2%), Coast (69.7%), and North Eastern province (74.0%).

Inequality in Kenya remains high. The distribution of income measured by the Gini coefficient was estimated at 39% in rural areas and 49% for urban areas (pre- crisis). Income disparities in the rural areas have gone down since 1997, while the disparities in the urban areas have increased slightly.

There has been additional progress with regard to other dimensions of social development over the past years. For example, net primary education enrolment was only 80% in 2003, but increased to about 90% in 2008 (with an equal enrollment ratio between boys and girls). In 2004, only about 60% of primary students completed their education compared with about 80% in 2008.

According to last Country Briefs,an estimated 3.8 million people in rural areas are between highly-to-extremely food insecure. FAO/ GIEWS and FEWSNET agree that, in the short term, Kenya is a hunger-prone country, while World Food Program and IFPRI assess the long-term situation as alarming and hunger as moderately high. There is a long history of periodic shortfalls in food supply in Kenya. Shortfalls occur in all the country or in parts of the country, and sometimes for two years in a row. In times of unfavourable weather, even the provinces normally characterised by a maize surplus (such as the Rift Valley) or marginally self-sufficient provinces (such as Western and Nyanza) may enter into maize deficit situation. In addition, in areas characterised by chronic deficits (such as the Coast and Eastern and North Eastern provinces) the situation becomes acute. In many districts in these areas, emergency relief becomes necessary.

A good poverty profile also makes employment targeting possible. The ability of the vast majority of households in Kenya to escape poverty will depend on their earnings from employment. The highest poverty rate was found among people living in households headed by farmers 46 % (KNBS, 2009). By contrast, households headed by someone working in the government are least likely to be poor; in these occupations the poverty rate was 20 % (1993–94). This would suggest that policies that aim to reduce poverty through enhancing income-generating capabilities should be targeted toward the agricultural sector.

The relationship between poverty and education is particularly important because of the key role played by education in raising economic growth and reducing poverty. The better educated have higher incomes and thus are much less likely to be poor. Kenyans living in households with an uneducated household head are more likely to be poor, with a poverty rate of 47 % in 2014, National Poverty Atlas (KNBS, 2007a).With higher levels of education, the likelihood of being poor falls considerably. Raising education attainment is clearly a high priority to improve living standards and reduce poverty.

The relationship between gender and poverty may also indicate another targeting strategy for poverty reduction. In Tanzania, about 35 percent of the population lives in households headed by women. Perhaps surprisingly, the 2007 data show that the poverty rate was slightly lower among female-headed households (48 %) than among male-headed households (52 %). In this case, targeting interventions based on the gender of the head of household would not help to distinguish the poor from the non-poor (Booysen *et al.*, 2008).

1.2 Defining and Measuring Poverty

Poverty is a worldwide concern. Although there is a global concern towards poverty reduction, there is a little agreement on a single definition and measurement of poverty (Kotler *et al.*, 2006; Laderchi *et al.*, 2003). According to Kotler *et al.* (2006) and Laderchi *et al.* (2003), the problem of arriving at one single definition of poverty has been compounded by a number of factors. Poverty affects heterogeneous groups such that the concept of poverty is relative depending on different interest groups and individuals experiencing it (Kotler *et al.*, 2006; Rank, 2004). The difficulty surrounding the definition and measurement of poverty has often led poverty researchers and policy makers to relate poverty to the concepts of impoverishment, deprivation, the disadvantaged, inequality, the underprivileged and the needy underscored four main approaches to poverty definition and measurement.

Monetary Poverty

According to Laderchi *et al.* (2003), the monetary approach defines poverty in terms of how much a person's income (or consumption) falls short of some minimum level of resources. The monetary approach to poverty measurement involves methodologies that emphasize monetary indicators and an objective derivation of the poverty line. The monetary approach is based on the assumption that a uniform monetary metric can be used to control for the heterogeneity of all the individuals and their situations. He further pointed out that determining poverty based on a monetary metric entails the choice of an indicator, a unit of analysis, and a poverty line.

A monetary indicator provides a common denominator of measurement for comparability . The dominant use of the monetary indicators to measure poverty is justified on the grounds that it can approximate aspects of poverty or well-being that are difficult to measure in the same unit. In addition, a monetary approach serves as a standard homogenous platform of poverty measurement that eases the tension between theoretical complexity and diversity of poverty definitions and measurements. The monetary approach emphasizes on the choice of income or expenditure indicator as a proxy for consumption as a proxy for permanent income. He also suggested a weakness of the monetary approach in its focus on the physical or moral character of the poor rather than the real causes of poverty.

Traditionally, poverty is viewed as an individual problem, even though many of the causes of poverty can be traced to the household level. Laderchi *et al.* (2003) suggested poverty analysis should consider the household as a unit of observation and the results of the analysis can be presented either at the household or individual level.

The choice of a poverty line is crucial to poverty measurement. A poverty line may be identified either with respect to a list of basic needs (absolute) or some characteristics of the distribution of the welfare indicators chosen (relative) (Laderchi *et al.*, 2003). Ravallion's Food Energy Intake method underscores the level of income or expenditure at which food energy requirements are met. The lack of economic theory to determine minimal level of needs caused the estimation of the poverty line to be influenced by political debates and policy agenda. Because the choice of poverty line has political influence and a lack of economic theory, the poverty line tends to be problematic and misleading (Laderchi *et al.*, 2003).

Determinants of Monetary Poverty

Monetary poverty is measured as the total income or consumption proxy by either expenditure or income. In most developing countries and the United States, the absolute poverty line is used and food energy requirements are taken into account for the development of the poverty line (Laderchi *et al.*, 2003).

The poverty threshold is computed by putting a monetary value on the minimum amount of food a family or individual needs to survive. When a family or an individual's total income falls below the poverty threshold, then the family or individual is considered poor. The family or an individual's monetary poverty level is associated with family size, age, gender, race, place of residence and marital status (Schiller, 2008; Hurst, 2004).

Schiller (2008) pointed out that an increase in family size has an important implication for family financial need and security. An increase in family size requires more demand for household services and goods such as an increase in family laundry and health care services. According to Schiller (2008), an increase in family size can be associated with an increased level of poverty. For instance, an increase in the number of children from one to five can triple the family poverty level. On the contrary, total family income is likely to increase with family size as more members of the family take up employment in the labor market.

Capability Poverty

Capability poverty is the failure of a person to achieve basic capabilities to adequately fulfill certain crucial functions at minimal level (Laderchi *et al.*, 2003; Sen, 1985). The capability approach views monetary resource as means that that can help to enhance people's well-being. The monetary resource is viewed as a necessary, but not sufficient condition to prevent the casual chain of poverty (Laderchi *et al.*, 2003). Therefore, the capability approach emphasizes both monetary resources and other resources to develop or achieve capabilities. Literature review on capability poverty primarily focused on the work of Sen (1985). Sen argued that the monetary approach emphasizes utility of a commodity and does not provide a good proxy to assess people's well-being.

Sen (1985)capability approach provides a framework that can be use to assess inequality, poverty and individuals' or groups' well-being. Sen's concept of capability operates at two levels: at the level of realized well-being or outcome measured by functioning, and at the level of potential well-being or opportunity measured by capability. Functioning refers to a person's achievement while capability refers to the combination of various functions a person can achieve. Sen, pointed out that a person's achievement or functioning is a better proxy for well-being. What a person successfully accomplished with a commodity is what matters, taking into consideration the characteristics of the commodity, the characteristics of the person and external circumstances.

Choosing an Indicator of Poverty

There are a number of ways to measure well-being. In the Welfarist approach, Sen (1979) seeks to measure household utility, which in turn is usually assumed to be approximated by household consumption expenditure or household income; these may be considered as inputs into generating utility. Given enough income, the household is assumed to know best how to deploy these resources, whether on food, clothing, housing, or the like. When divided by the number of household members, this gives a per capita measure of consumption expenditure or income. Of course, even household expenditure or income is an imperfect proxy for utility; for instance, it excludes potentially important contributors to utility such publicly provided goods or leisure.

If we choose to assess poverty based on household consumption or expenditure per capita, it is helpful to think in terms of an expenditure function, which shows the minimum expense required to meet a given level of utility μ , which is derived from a vector of goods X, at prices p. It can be obtained from an optimization problem in which the objective function (expenditure) is minimized subject to a set level of utility, in a framework where prices are fixed (Chen, 2007).

Let the consumption measure for the household i be denoted by y_i. Then an expenditure measure of welfare may be denoted by:

$$y_i = p.qe(p,x,\mu) \tag{1.1}$$

where p is a vector of prices of goods and services, q is a vector of quantities of goods and services consumed, $e(.)$ is an expenditure function, X is a vector of household characteristics (number of adults, number of young children, and so on), and μ is the level of "utility" or well-being achieved by the household. Put another way, given the prices (p) that it faces, and its demographic characteristics (x), y_i, it measures the spending that is needed to reach utility level μ.

Typically, we compute the actual level of y_i from household survey data that include information on consumption. Once we have computed y_i we can construct per capita household consumption for every individual in the household, which implicitly assumes that consumption is shared equally among household members. For this approach to make sense, we must also assume that all individuals in the household have the same needs. This is a strong assumption, for in reality, different individuals have different needs based on their individual characteristics.

Other possible measures of well-being include the following:

- Calories consumed per person per day. If one accepts the (Non-Welfarist) notion that adequate nutrition is a prerequisite for a decent level of well-being, then we could just look at the quantity of calories consumed per person. Anyone consuming less than a reasonable minimum often set by World Health Organization at 2,100 Calories per person per day would be considered poor. However, at this point we just note that it is not always easy to measure calorie intake, particularly if one wants to distinguish between different members of a given household. Nor is it easy to establish the appropriate minimum number of calories per person, as this will depend on the age, gender, and working activities of the individual.

- Food consumption as a fraction of total expenditure. Over a century ago, Ernst Engel noted that in Germany that as household income per capita rises, spending on food rises too, but less quickly. This relationship is shown in Figure 1.1. As a result, the proportion of expenditure devoted to food falls as per capita income rises. One could use this finding, which is quite easy to come up with as a measure of well-being and hence as measure of poverty. For instance, households that devote more than (say) 60 percent of their expenditures on food might be considered to be poor. The main problem with this measure is that the share of spending going to food also depends on the proportion of young to old family members (more children indicates a higher proportion of spending on food), and on the relative price of food (if food is relatively expensive, the proportion of spending going to food will tend to be higher) (Chen, 2007).

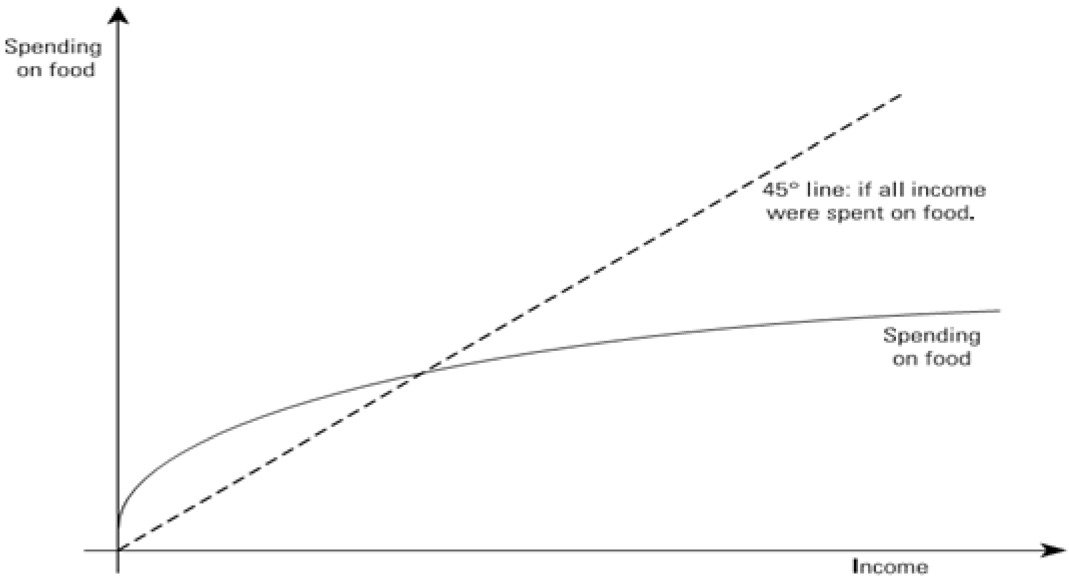

Figure 1.1: Engel Curve: Food Spending Rises Less Quickly Than Income

1.3 Measures of Poverty

This study seeks to come up with measures of poverty for policy mitigation in the Lake Victoria basin which from the country brief survey has high levels of poverty

A poverty measure is an index that shows the magnitude of poverty in a society. To form such a measure, an aggregation formula is required that sums up the income dimensions of poverty for a given population (GoK, 1998; Mwabu *et al.*, 2000). One poverty measure that has been found manageable in presenting information on the poor in an operationally convenient manner is the Foster, Geer and Thorbecke (FGT) measure developed by Foster *et al.* (1984). This measure is used to quantify the three well known elements of poverty: the level, depth and severity (also known as incidence, inequality and intensity, respectively) of poverty. The FGT formula that is normally used to measure overall income poverty is shown in the Equation 1.2.

$$P_\alpha = \frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{Y_i}{z}\right)^{\alpha} \tag{1.2}$$

Where:

P_α is a measure of absolute poverty, including food poverty,

Y_i is the total expenditure of household i, expressed in per adult equivalent ($i = 1, ..., N$),

Z is the poverty line, expressed in per adult equivalent,

N is the total number of households, and

α is the FGT parameter, which may be interpreted as a measure of poverty aversion, $\alpha \geq 0$.

Inequality refers to the variations in living standards or well being across a whole population. It is the fundamental disparity that permits one individual certain material choices while denying another those very same choices. The Gini coefficient is generally used to measure levels of inequality. It takes on values between 0 and 1 with zero interpreted as no equality. Graphically, the area between the Lorenz curve and the line of equality can easily represent the Gini coefficient (Kuznets, 1955). This is illustrated in Figure 1.2.

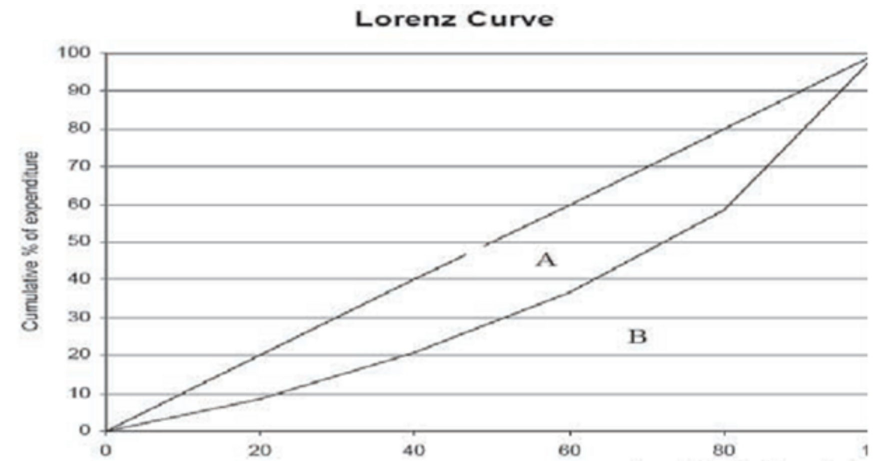

Figure 1.2: An illustration of Lorenz Curve and the GINI coefficient

The Gini coefficient is calculated as the area A divided by the sum of areas A and B. If income is distributed completely equally, then the Lorenz curve and the line of total equality are merged and the Gini coefficient is zero.

Given information on a welfare measure such as per capita consumption, and a poverty line, the next issue is deciding on an appropriate summary measure of aggregate poverty. A number of aggregate measures of poverty can be computed. The formulae presented in this section are all based on the assumption that the survey represents a simple random sample of the population, which makes them relatively easy to understand. Some of commonly used measures of poverty indices are described below in the study.

1.3.1 Headcount Index

By far, the most widely used measure is the headcount index, which simply measures the proportion of the population that is counted as poor, often denoted by P_0.

The greatest virtues of the headcount index are that it is simple to construct and easy to understand. These are important qualities. However, the measure has at least three weaknesses: First, the headcount index does not take the intensity of poverty into account. Second, the headcount index does not indicate how poor the poor are, and hence does not change if people below the poverty line become poorer.

Table 1.1: Headcount Poverty Rate in A and B assuming Line of 125

Countries	Expenditure for each individual country				Headcount poverty level p_o
Country A	100	100	150	150	50%
Country B	124	124	150	150	50%

clearly there is greater poverty in country A, but the headcount index does not capture this. As a welfare function, the headcount index is unsatisfactory in that it violates the transfer principle (Ravallion, 1996a).

1.3.2 Poverty Gap Index

This measure is the mean proportionate poverty gap in the population (where the non-poor have zero poverty gap). Some people find it helpful to think of this measure as the minimum cost of eliminating poverty (relative to the poverty line), because it shows how much would have to be transferred to the poor to bring their incomes or expenditures up to the poverty line (as a proportion of the poverty line). The minimum cost of eliminating poverty using targeted transfers is simply the sum of all the poverty gaps in a population; every gap is filled up to the poverty line. However, this interpretation is only reasonable if the transfers could be made perfectly efficiently, for instance, with lump sum transfers, which is implausible. Clearly, this assumes that the policy maker has a lot of information; one should not be surprised. To find that a very "pro-poor" government would need to spend far more than this in the name of poverty reduction.

Table 1.2: Poverty Gap Index, assuming Poverty Line of 125

Country	Expenditure for each individual country				Poverty gap index, p_1
Country C	100	110	150	160	
Poverty gap	25	15	0	0	
$G_{i/z}$	0.20	0.12	0	0	0.08 = 0.32/4

Thus this measure is an indicator of the potential saving to the poverty alleviation budget from targeting ; the smaller is the poverty gap ,the greater the potential economies for a poverty alleviation budget from identifying the characteristic of the poor-using surveys or other information-so as to target benefits and programs.

1.3.3 Squared Poverty Gap (Poverty Severity) Index

This is simply a weighted sum of poverty gaps (as a proportion of the poverty line), where the weights are the proportionate poverty gaps themselves. The measure implicitly puts more weight on observations that fall well below the poverty line.

Table 1.3: Squared Poverty Gap Index, assuming Poverty Line of 125

Country	Expenditure for each individual country				Squared Poverty gap index, p_2
Country C	100	110	150	160	
Poverty gap	25	15	0	0	
$G_{i/z}$	0.20	0.12	0	0	
$\left(G_{i/z}\right)^2$	0.04	0.0144	0	0	0.0136 = 0.0514/4

An important aspect in poverty analysis is the setting of poverty lines below which persons are considered to be poor and above which they are not poor. The food component of the poverty line is divided by some estimate of the budget share devoted to food to obtain the overall poverty line. The non food component is then got by taking the difference between the overall and the food poverty lines. The problem here is that the determination of the budget share devoted to food is likely not to be a transparent process (Ravallion, 1998). Due to the intrinsic weaknesses in the FGT measures, the Food Energy Intake (FEI) and CBN measures and in general all summary measures of food poverty, what is ideally needed is an approach that is good as the choice of food poverty line or measure.

There is a lack of consensus on how to measure poverty in general, even though poverty indices and poverty profiles are increasingly being used as guides in targeting resources to reduce poverty. An allocation that is efficient according to one methodology may yield unacceptable results when a different methodology is applied.

Results from poverty studies are also sensitive to the choice of poverty line (the means of identifying the poor) and poverty measure (the measure obtained when aggregating incomes or expenditures of households below poverty line). Since the choices are typically at the discretion of the analyst, this has given rise to the suggestion that the results obtained are not robust. Potentially different results could be obtained by the choice of a different poverty line or measure. Moreover, few conclusions can be drawn if poverty trends differ substantially when different poverty measures are applied or the position of the poverty line is changed.

Analysts have tried to overcome the problem by employing a number of poverty lines/measures but this only partially overcomes the problem since it may still be possible to obtain different results by the choice of another poverty line. Thus, what is ideally needed is an approach that is robust to the choice of poverty line.

Most of the studies on poverty in Kenya (Kabubo-Mariara, 2007; Geda *et al.*, 2001; Oyugi, 2000) have used the summary measures to determine the extent and level. The robustness of poverty measures using summary measures such as means and variances can be compromised by errors in living standards data, unknown differences between households at similar consumption levels, uncertainties and arbitrariness in both the poverty lines and the precise poverty measure.

Reducing poverty and improving household food security is an important policy instrument for the development in Africa.Many pro-poor development programs have been introduced over the past decade to bring the cycle of poverty and food insecurity to an end.

1.4 Food crop Balance sheets

This chapter aims to investigate the empirical analysis of food balance sheet in the region to examine the policy needs to be implemented for the intervention against food insecurity at regional and household levels respectively.

Food balance sheet (FBS) presents a comprehensive picture of the pattern of the country's food supply during a specified reference period. The food balance sheet shows for each food item that is, each primary commodity and a number of processed commodities potentially available for human consumption, the sources of supply and its utilization. The total quantity of foodstuffs produced in a country added to the total quantity imported and adjusted to any change in stocks that may have occurred since the beginning of the reference period gives the supply available during that period. on utilization side a distinction is made between the quantities exported,fed to livestock ,used for seed ,put to manufactured for food use and other uses, losses during storage and transportation and food supplies available for human consumption. The per capita supply of each such food item available for human consumption is then obtained by diving the respective quantity by the related data on the population actually partaking of it. Data on per capita food supplies are expressed in terms of quantity and by applying appropriate food composition factors for all primary and processed products also in terms of caloric value and protein and fat content.

Annual food balance sheet tabulated regularly over a period of years will show the trends in the overall national food supply, disclose changes that may have taken place in the types of food consumed, that is, the pattern of the diet and reveal the extent to which the food supply of the country, as a whole, is adequate in relation to nutritional requirements (See Appendix 2).

By bringing together the larger part of the food and agricultural data in each country, FBS also can serve in the detailed examination and appraisal of the food and agricultural situation in a country. A comparison of the quantity of food available for human consumption with those imported will indicate the extent to which a country depends upon imports (import dependency ratio). The amount of food crops used for feeding livestock in relation to total crop production indicates the degree to which primary food resources are used to produce animal feed which is useful to know when analyzing livestock policies or patterns of agriculture. Data on per capita food supplies serve as a major element for the projection of food demand, together with other elements, such as income elasticity coefficients, projections of private consumption expenditure and of population.

It is important to note that the quantities of food available for human consumption, as estimated in the FBS, relate simply to the quantities of food reaching the consumer.

However, the amount of food actually consumed may be lower than the quantity shown in the balance sheet depending on the degree of losses of edible food and nutrients in the households, e.g. during storage, in preparation and cooking (which affects vitamins and minerals to a greater extent than they do calories, proteins and fat), as plate-waste or quantities fed to domestic animals and pets, or thrown away.

Food balance sheets do not give any indication of the differences that may exist in the diet consumed by different population groups, e.g. different socioeconomic groups, ecological zone and geographical areas within a country, nor do they provide information on seasonal variations in the total food supply. This study will be considering developing a mathematical model for forecasting per capita food intake in the Lake Victoria basin.

1.5 Statement of the problem

According to the WB (2001) and Chen and Ravallion (2002), poverty policies have utilized a broad conceptualization of poverty associated with different dimensions of poverty. Schiller (2008), Laderchi *et al.* (2003) and Jordan (1996) pointed out that the way we conceptualize and measure poverty influences the fundamentals of poverty policies and programs. While different poverty measures have been utilized, little attention has been paid to their comparative outcomes and implications (Bell, 1995; Schiller, 2008).

Laderchi *et al.* (2003) and Hagenaars and Vos (1988) emphasized that the choice of a specific definition and measurement of poverty may result in different estimates of the determinants of poverty and evaluation outcomes for poverty programs. However, researchers and policymakers often prefer to adopt a particular definition of poverty based on the availability of data, political interest or historical justification. While the choice of a specific poverty indicator may have major consequences for poverty reduction, some indicators may be a better measure for a specific poverty situation (Hagenaars and Vos, 1988; Laderchi *et al.*, 2003).

The official poverty measure is noted to have both methodological and resource definition flaws (Dalaker and Naifeh, 2005). A poverty threshold based on a simple commodity is inappropriate because it makes the threshold numbers more sensitive to the price of that food than the price of any other expenditure for low-income families. While many poverty studies utilize the world Bank poverty threshold, other evaluative studies tend to focus on indicators of capability and social exclusion poverty (Rank, 2004). The different dimensions of poverty add to the problem of choosing the appropriate poverty measure and indicators. What is the appropriate measure to estimate the incidence of poverty? In other words, what criteria of poverty should be used to define and measure poverty? What is missing from previous studies is an analysis of different poverty measures in lake Victoria basin.

1.6 Research Objectives

The study was guided by the following objective.

1.6.1 General Objective

The general objective was to develop a logistic model and an augmented model that can be used to reliably assess poverty profiles and food insecurity in the Lake Victoria basin of Kenya.

1.6.2 Specific Objectives

1. To propose a method of measuring poverty by defining an indicator of welfare and a minimum acceptable standard of the indicator.
2. To investigate the performance of the logistic and the augmented in the determination of poverty predictors in the Lake Victoria Basin.
3. To develop a mathematical model for forecasting food crops balance sheet as a tool for early warning system in the Lake Victoria Basin.

1.7 Organization of the thesis

This thesis is organized into six chapters as follows the first chapter contains the background, the statement of the problem, objectives of the study, significance of the study, research hypothesis and the scope of the study. The second chapter deals with the review of related literature. It includes the concept of food security, the food Balance sheet and the components of the food balance sheet and their definition. The third chapter presents a detailed account go the methodology used to accomplish the research objectives. This includes the study area, sources of data and the acquisition methods. It emphasizes the logistic regression theoretical framework. Chapter four presents the estimation methods of the model parameters, the study checks the consistency and asymptotic properties of the parameters. Chapter five presents the main finding of the research, the results of household survey on determination of food security and the household perceptions. Chapter six concludes by presenting the issues discussed in this study by providing recommendation for further improvement.

CHAPTER TWO

LITERATURE REVIEW

The aim of this chapter is to review the available literature related to poverty,poverty profiles and the FBS. A critical analysis of what other researchers have said on the subject is presented in this chapter. It targets definitions of poverty and the poverty profiles literature and the mathematical models within FBS as a tool to aid decision of county government policy strategies on food security. This section describes the previous poverty studies in Kenya and around the world ,also the relationship between FBS and individual food intake and dietary diversity. The section also suggests various method used to measure poverty and areas FBS has been applied in literature.

2.1 Poverty line models

A poverty line may serve other purposes such as monitoring poverty over time, developing poverty profiles, acting as a threshold for entitlement and providing a focus for public debate (Ravallion and Bidani, 1994)

An absolute poverty line has fixed real values over time and space while a relative poverty line has values that rise with average expenditure. Ravallion (1998) argues that a poverty line should always be absolute in the space of welfare for purposes of informing anti-poverty policies. Such a poverty line guarantees that the poverty comparisons made are consistent in the sense that two persons with the same level of welfare are treated the same way.

The traditional techniques for constructing poverty lines are the FEI and the Cost of Basic Needs (CBN) methods. Both methods anchor the definition of basic needs to food energy requirements.

The FEI aims at finding a monetary value equivalent at which basic needs are met. It presents only the minimum level of basic needs, below which a material lifestyle is not possible. The FEI method sets the minimum food requirement by setting the consumption expenditure level at which food energy intake is just sufficient to meet predetermined average food energy requirement for normal body function. This approach has been widely used by various authors such as Foster *et al.* (1984); Ercelawn (1991) and Ravallion and Bidani (1994). The advantage with the FEI method is that it automatically includes an allowance for both food and non-food consumption thus avoiding the tricky issue of determining exactly the basic needs of these goods as long as one locates the total consumption expenditure at which a person typically attains the calorie requirement. It also does not rely on price data which can be a problem in many developing countries. It is also parsimonious in its data requirements and it allows for differences between subgroups (Madden, 2000).

The FEI has inherent weaknesses at the basis for welfare comparisons. The FEI poverty line is computed under the strong assumption that food expenditure and calorie intake are not independently observed (Bouis and Haddad, 1990; Madden, 2000). As noted by Greer and Thorbecke (1986b), the use of fixed food weight to calorie intake factor for the whole country over time and over entire income profile might be inappropriate due to changing food quality and food preparation methods. The method does not allow us to make comparisons across different subgroups of the population using a common yardstick for standards of living because it suffers from inconsistency problems Ravallion and Bidani (1994). The relationship between food energy intake and total consumption is likely to differ according to differences in tastes, activity levels, relative prices or publicly provided goods.

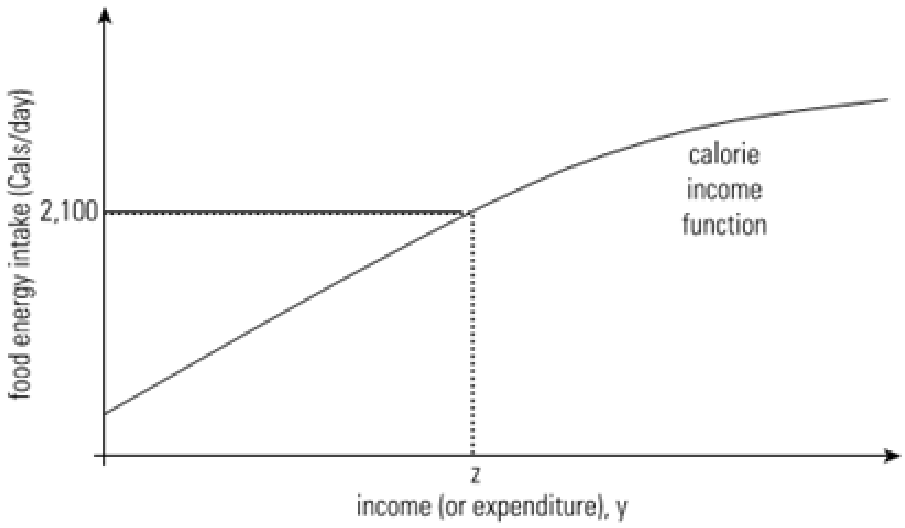

Figure 2.1: Calorie Income Function

In the CBN approach, the poverty line is constructed by determining a food basket which is assumed to be bought by all households (KNBS, 2007a). This approach considers poverty as a lack of command over basic consumption needs and the poverty line as the cost of those needs. The basic food basket is usually set using the nutritional requirements. The bundle is then evaluated at local prices to get the food component of the overall poverty line

In the measurement of household poverty, a standard adequate food basket was established and compared with the actual food basket of a given household. Two approaches were used to derive a household food basket: The first, i.e., the Least Cost Approach derived food basket which met the minimum caloric requirements at lowest cost given prevailing market prices. This approach did not consider preferences of the poor; it instead assumed they purchased the cheapest foods. This was not always the case. The second was the Expenditure – based approach suggested by Lanjouw and Ravallion (1995) (www.undp.org) referred to by Malaba (2006) which was adopted for the Zimbabwe experience in which they looked at expenditure patterns of the bottom 40 households. The food items the households ate were listed. All monthly expenditure was converted to one reference period using CPI. Food items from purchases, transfers and own production were then weighted using expenditure shares and quantities. The first 30 items with the highest weighted expenditure formed the food basket. Average monthly food expenditure per capita & number of grams per capita per month were computed using prevailing commodity prices collected in a survey and weighted using quantities and values households reported to have produced, received as transfers or purchased. For each item, a weighted mean price was used. Grams per capita were converted to calorific values per capita per day. Total calories consumed per capita per day from all the food items in the food basket were determined. Mean monthly expenditure on them were also determined. The Food Security threshold or Food Poverty Line (FPL) was computed as:

$$FPL = \frac{\text{Total carolies consumed}}{\text{2100 kilo calories}} \times \text{mean monthly expenditure} \tag{2.1}$$

2.1 Small Area Estimation Technique: Household-level Method

This was a multivariate regression models which used a small area estimation technique referred to as household-level method documented by Hentschel *et al.* (2000); Elbers *et al.* (2001). It required a minimum of two sets of data: household-level census data and a representative household survey corresponding approximately to the same period as the census. The first step was to estimate a model of consumption-based household welfare using household survey data with explanatory variables limited to those found in both data sets.

The equation

$$lnC = \alpha + \beta_1 X + \varepsilon \tag{2.2}$$

was estimated using ordinary least squares, where C was total per-capita consumption, or another poverty proxy, X a matrix of household-level characteristics . The resulting parameter estimates were applied to the census data. For each household, the estimated parameters from the regression were used to compute the probability of each household in the census living in poverty. The household-level value of the explanatory variable was multiplied by the corresponding parameter estimate. The estimated value of the benchmark indicator was then used to determine the probability of a household being food-insecure or poor in terms of given threshold below which a household was non-poor or poor whether based on consumption, caloric intake or anthropometric measures.

Here:

$$\begin{aligned} F_{ij} &= 1 \text{ if } lnC_{ij} < lnz; \\ \text{and } F_{ij} &= 0 \text{ otherwise} \end{aligned} \tag{2.3}$$

Following Hentschel *et al.* (2000) and using the model of consumption from equation but with only one vector of explanatory variables, the expected poverty status of household i was obtained as:

$$E\left[F_{ij}|X_i, \beta, \alpha\right] = \Phi\left(\frac{lnz - X_i\beta}{\sigma}\right) \tag{2.4}$$

Where: Φ was the cumulative standard normal distribution. This equation gave the probability that a household was poor. The challenge with this method was that it required two sets of data which should have been collected during the same periods of time. Getting variables which match in the two data sets could be a challenge unless they are planned together with the intention of using the two for the purpose of small area estimation. This study proposed the use of one survey dataset for the model which could be empirically obtained or obtained from previous surveys.

2.1.1 Previous Poverty Studies in Kenya

The Kenyan economy was regarded as an African success story early into the post-independence years of many African countries. In the 1960s and 1970s, the country achieved a high growth rate of 6.6 per cent per annum. However, this rapid rate of growth was not sustained thereafter. Between 1974 and 1979, the growth rate declined to 5.2 per cent per annum. Further declines occurred in the 1980-89 and 1990-95 periods when the average growth rates averaged 4.1 and 2.5 per cent per annum respectively. Over the plan period 1997-2001, the target was set at 5.9 per cent per annum. However, contrary to expectations, the economy registered a negative growth rate of 0.3 per cent in the year 2000. The decline was reflected in almost all the sectors of the economy. The GDP per capita was estimated at US $ 275 in 1995 and stood at US$ 294 in 2000. Because of this poor economic performance, about 13.6 million Kenyans in 2000 lived under the poverty line, and the situation has continued to worsen. In the context of growing inequalities, and increasing absolute poverty in rural and urban areas, there is need to understand regional and institutional factors associated with poverty. Though a large number of studies now exist on Kenyan poverty, its measurement and determinants Greer and Thorbecke (1986b); Mukui (1994); Mwabu *et al.* (2000); Oyugi (2000); Mwabu *et al.* (2002); Geda *et al.* (2005), there is a deft of empirical studies on institutional determinants of poverty in Kenya. Oiro *et al.* (2004) only employ descriptive methods to explain the impact of rural institutions on poverty. This study is a response to this research gap. We build on the existing studies on determinants of poverty and Oiro *et al.* (2004) to analyze the institutional perspectives of poverty.

Analytical work on determinants of poverty in Kenya is at best scanty. Most of the available studies are descriptive and focus mainly on measurement issues. Earlier poverty studies have focuses on a discussion of inequality and welfare based on limited house level data. One recent comprehensive study on the subject is that of Mwabu *et al.* (2000), which deals with measurement, profile and determinants of poverty. The study employs a household welfare function, approximated by household expenditure per adult equivalent. The authors runs two categories of regression, using overall expenditures and food expenditures as dependent variables. In each two cases, three equations are estimated which differ by type of dependent variable. The dependent variables are: total household expenditure, total household expenditure gap (the difference between the absolute poverty line and the actual expenditure) and square of the latter. A similar set of dependent variables is used for food expenditure, with explanatory variables being identical in all cases.

Geda *et al.* (2001)justified their choice of this approach (compared to a logit/probit model) as follows. First, the two approaches (discrete and continuous choicebased regressions) yield basically similar results the logit/probit model involves unnecessary loss of information in transforming household expenditure into binary variables. Although their specification is simple and easy to follow, it has certain inherent weaknesses. One obvious weakness is that, unlike the logit/probit model, the levels regression does not directly yield a probabilistic statement about poverty. Second, the major assumption of the welfare function approach is that consumption expenditures are negatively associated with absolute poverty at all expenditure levels. Thus, factors that increase consumption expenditure reduce poverty.However, this basic assumption needs to be taken cautiously. For instance, though increasing welfare, raising the level of consumption expenditure of households that are already above the poverty line does not affect the poverty level (as for example measured by the headcount ratio)

Notwithstanding such weakness, the approach is widely used, Mwabu *et al.* (2000) identified the following as important determinants of poverty: unobserved region-specific factors, mean age, size of household, place of residence (rural versus urban), level of schooling, livestock holding and sanitary conditions. The importance of these variables does not change whether the total expenditure, the expenditure gap or the square of the gap is taken as the dependant variable. The only noticeable change is that the sizes of the estimated coefficients are enormously reduced in the expenditure gap and in the square of the expenditure gap specifications. Moreover, except for the minor changes in the relative importance of some of the variable, the pattern of coefficient again fundamentally remains unchanged when the regressions are run with food expenditures as dependant variable.

Another recent study on the determinant of poverty is Oyugi (2000), which is extension to earlier work by Greer and Thorbecke (1986b,a). The latter study used household calorie consumption as the dependant variable and a limited number of household characteristics as explanatory variables. An important aspect of Oyugi (2000) study is that it analyse poverty both at micro (household) and meso (district) level, with the meso-level analysis being the innovative component of the study. Oyugi (2000) estimate a probit model using data of the 1994 Welfare Monitoring Survey data. The explanatory variable (household characteristics) include: holding area livestock unit, the proportion of household members able to read and write, household size, sector of economic activity (agriculture, manufacturing/industrial sector or results of the probit analysis show that all variable used are important determinants of poverty in rural areas and at the national level, but that there are important exceptions for urban areas Oyugi (2000). These results are consistent with those obtained from the meso-level regression analysis.

It is interesting to compare the implications of the logit model used by Mwabu *et al.* (2000) and probit model used Oyugi (2000) regression approaches. In the probit model, in order of importance the key determinants of poverty are: being able to read and write, employment in off-farm activities, being engages in agriculture, having a side-business in the service sector, source of water and household size. Region of residence appears to be equally important in determining poverty status in the two approaches. Although the two approaches did not employ the same explanatory variables, this comparison points to the possibility of arriving different policy conclusions from the two approaches.

2.2 Food Balance Sheet Studies

The FAO suggests various uses of FBS, but they also caution that the estimates for national food or nutrient availability do not deal with distribution of food or nutrient supply between regions within country or among other groups of household. FAO suggests that the data may be used to;

1. Observe a country's food supply and trends

2. Compare food supply with nutritional requirement for healthy diets
3. Estimate supply/ shortage measures
4. Evaluate food and nutritional policies
5. Investigate relationships between food supplies ,famine, and malnutrition
6. Set goals for trade and production and project future supply and demand

FBS has been used as a measure of undernourishment. These measures are used by policy makers, planners and non-governmental organizations to direct resources to address nutritional concerns (Smith, 2009). Smith (2009) and Svedberg (2011) detailed how FAO uses FBS figures for daily per person caloric value as the mean of lognormal distribution of each country's caloric availability from which it determines the country's probability of not meeting a minimum dietary energy requirement. The spread of the distribution is determined by the variability in dietary intake over the country's population that is determined by a household survey. In the "Handbook for the Preparation of Food Balance Sheet" the FAO provides several cautionary notes regarding FBS data in general. Data related to changes in stocks are of particular concern, due to lack of complete and quality data. FAO notes that variability in stock changes is a main motivation for the publication of the FBS as three year moving averages. However, the FBS are also provided as annual time series that are regularly updated and revised which can be manipulated to create averages over any number of years.

Since many numbers given in the FBS are estimates, it is important to know the degree of the potential error. Figures for food estimates and stock changes are believed to be subject to considerable potential error (Gillin, 2000). As a result, the estimates of food supply derived from subtracting feed and stock changes would also contain substantial error. Svedberg (2012) assessed the sensitivity to potential errors by the following methodology. For Sub-Saharan Africa, Svedberg (2012) followed the FAO procedure for estimating the percentage of the population that is undernourished. He then introduced a 10% error in either direction in the estimate of daily calorie intake per person. He used the FAO's 1800 kcal/day individual minimum cutoff point and the FAO-estimated coefficient of variation, the result was a very large variation in the share of the population that was undernourished. Svedberg (2012) exercise shows that data errors well within plausible range have considerable effects on the estimates of the share of undernourished populations.

The most common use of FBS data in published literature is the citation of daily energy intake and fat and protein intake (Grigg, 1993, 1996). Estimates of intakes of other nutrients include; vitamins, minerals, and amino acids are also based on FBS data on food availability. Also, trends and changes over time for the intake of energy and various nutrients are examined using FBS. In this study, we will examine, using regression model, the relationship between trends and change over time with mean per capita calories.

FBS data is also used to compare food availability over time and among countries (Grigg, 1993, 1996; El-Obeid *et al.*, 1999; Diaz-Bonilla *et al.*, 2003). For instance Hopper compares per capita daily intake of calories in India, Japan and China during the period of 1955-1995. Grigg (1996) compares percentage of calories derived from starchy staples in developed and less developed countries between two time periods. In order to facilitate such comparisons over time, the FBS data is regularly revised for past years to the extent that errors are similar over time or across similar regions, comparisons may be more accurate and useful than the actual level of food availability. In this study we compare the capita intake in the three areas of the study.

The income measure is a categorical variable converted to a continuous variable; in-kind income from food assistance programmes such as Food Stamps is not included in the income measure and the food expenditure variable does not include in-kind programme benefits like Food Stamps or home-grown food.

In a study conducted in USA, the 18-question module provided a means of measuring both the prevalence of food security and the severity of hunger in the United States. Validation of the food security scale found that food insecurity is significantly negatively correlated with income and household food expenditures. The qualitative food security scale also correlated significantly with the more traditional measures, such as energy intake per capita (Laderchi *et al.*, 2003).

FAO publishes updated calculations of per capita food availability derived from food balance sheets and based on national averages. The average per capita availability in developing countries is lower than that in developed market economies. The former is estimated to provide 93% of the defined energy requirements and 115% of the requirements for the latter (FAO, 1985). In essence, the national average per capita calorie intake determines the number or the proportion of under nutrition people for each country. Although the FAO and the World Bank have attempted to take account of some personal characteristics in defining undernutrition, undernutrition has been considered as a proxy for undernourishment in the literature. In addition, other objections have been raised to reflect the extent and the scope of undernourishment. Consequently, the philosophy and dimensions of food security have been influenced by Sen (1981) concept of entitlement.

In this long term equilibrium view, food security is defined as access by all people at all times to enough food for an active healthy life (Sen, 1981; WB, 1986). The essential elements of this definition are the emphasis on both the demand (access) and the supply (availability) of food. Hence, food insecurity is simply the lack of access to enough food. The elaboration on the definition and the underlying conceptual framework of this approach focus attention on issues ignored in the previous definitions. These are the distinctions between transitory and chronic insecurity, inequalities in the distribution of income and wealth, seasonality and inter-annual variation and the functionality of an adequate diet. Utilizing data from 42 studies in non-representative population samples in Asian and Sub-Saharan African countries that provide figures on prevalence of mean per capita calories for both males and females, analysis of variance indicates that per capita calorific availability were low for woman than for man (FAO, 2002). In view of these findings, it appears that a number of issues need further investigation. The proposition that female deprivation is a major causative factor of high levels Household Food Insecurity (HFI) will be checked by incorporating this variable among per capita calorific determinants variables to check its significance.

Food production is one of the principal components of FBS. The concern of this study is to investigate whether there exists a relationship/correlation between food production and the per capita food availability.

Food security in general is a concept that integrates a number of important issues the magnitude of which ranges from micro to macro-economics. Its attainment requires no overall consideration in terms of policy and program development in all aspects of the food system. Hence, the success in production and distribution plays an important role in influencing the food security status of an individual, a household or a society at large (Maxwell and Smith, 1992). Food security is dependent on the ability of a population to access food in quantities and qualities that satisfy the dietary needs of individuals and households through the year.

The conceptual frame work of food security has progressively developed and expanded along with the growing incidence of hunger, famine and malnutrition in developing countries. In the mid-1970s food security was conceived a adequacy of food supply at global and national levels. This view focused merely on food production variables and overlooked the multiple forces that in many ways affected food access. In the 1980's the concept of food security attained wider attention that shifted from global and national level to household and individual levels. An understanding of food security also includes the time dimension, which explicitly describes the intensity and characteristics of household's food insecurity. Food insecurity can be "chronic" or "transitory". A contact failure to "access" food is distinguished as chronic, while a temporary decline is considered as transitory food insecurity. Chronic food insecurity is a sign of poverty and shows a long-term structural deficit in food production and lack of purchasing power. Transitory food insecurity, on the other hand, implies a short-term availability of food prices, production and income (Maxwell and Smith, 1992). Transitory food insecurity is a temporal or seasonal shortage of food because of unexpected factors for only a limited period and it I often triggered by seasonal instability in food supply or availability and fluctuation in prices and incomes. Chronic food insecurity can translate into a higher degree of vulnerability to famine or hunger. Repeated seasonal food insecurity also deplete the assets of the households and expose them to a higher level of vulnerability.

The World Bank defines food security as, "year round access to the amount and variety of food required by all household members in order to lead active and healthy lives, without undue risk of losing such access" (WB, 1986). This definition also encompasses availability, access and utilization to meet an active and healthy life. Household for security is the application of this concept at family level, with individuals in the household as the focus of concern. This suggests that, an analyst of household food insecurity should also focus on individual household members, i.e individual level of security within a household or the vulnerability of certain groups of a population due to their social status, labour availability and special nutritional needs such as rural women, malnourished children and the elderly. In some societies for instance, traditional or cultural practices prevent children and women to share the available food with men. Women may have less control of resources than men. Hence, women and children may be more vulnerable.

The World Food Summit 1996, defines food security as: "Food security exists when all people at all times, have physical and economical access to sufficient, safe and nutritious food to meet their dietary needs and food performances for an active, healthy life" (FAO, 1996). FAO has defined food security not in terms of access to, and availability of food, but also in terms of resources distribution to produce food and purchasing power to buy food, where it is produced.

FIVIMS, similarly, defines food security as a state that exists when all people, at all time, have physical, social and economic access to sufficient, safe and nutritious food which meets their dietary needs and food preferences for an active life. Food insecurity is due to unavailability of food, insufficient purchasing power, inappropriate distribution, or inadequate utilization at household level. Besides, vulnerability is also seen to be key, referring to factors that place at risk of becoming food insecure or reducing their ability to cope Hurst (2004). Moreover, food insecurity is a complex phenomenon attributable to a range of temporary and spatially varying vulnerability factors such as the socio-economic and political environment, the performance of the food economy, care practice and the health and sanitation situation (FAO, 2003). These are taken as indicators and key vulnerability factors that causes hunger and that should be monitored in assessing food insecurity (Ravallion, 1998).

2.3 Food Security Components

Food security is a multi-dimensional having interrelationships with vulnerability indicators; it cannot be captured by any single or specific indicator. It would therefore be important to understand the essential dimensions of food security - access to food, availability of food, and utilization of food. The interactions and combination of these dimensions represents food security together. Currently stability is also considered as the fourth component of food security (GTZ, 2006).

Access is referred to access by individuals to adequate resources (entitlements) to acquire appropriate foods for a nutritious diet. Entitlements are defined as the set of all those commodities bundles of a person can establish command given the legal, political, economic and socio arrangements of the community in which he/she lives (Including traditional rights - e.g access to common resources). Securing access to enough food at all times for an acting unhealthy live is a crime objective of all modern society because of the role prayed by food in economy, culture, and politics. Food access is largely determined by the ability of households and individuals to obtain food from own production, purchases and other sources, such as gifts, government transfers and food aid.

Availability referrers to the availability of sufficient quantities of food of appropriate qualities, supplied through domestic production or imports (including food aid). On the supply side cereal out put is the key indicator, as cereals provide about 60% on dietary energy in developing countries. At micro or household level, availability is taken as the capacity of the households to produce the food they need.

Utilization is related to utilization of food through adequate diet, clean water, sanitation, and the health care, to reach a state of nutritional well being for which all physiological needs are met. These brings out the importance of non-food inputs in food security. Its not enough that someone is getting what appears to be an adequate quantity of food if that person is unable to make use of the food because he/she is often falling sick. The dimension of food utilization underlines the importance of such process, including marketing, storage, processing, cooking practices, feeding practices and nutrition to the attainment of food security.

Stability is a very important component of the food security indicator. To be food secure a population, household or individual must have access to adequate food at all times. They should not be at risk of losing access to food as a consequence of a shock e.g.

an economic or climatic crisis or cyclically e.g. during a particular period of the year - seasonal food insecurity. The concept of stability can therefore refer to both the availability and access dimensions of food security.

Figure 2.2: Food Security Components

For example, food availability may be constrained by inappropriate agricultural knowledge, technology, policies, inadequate agricultural knowledge inputs, family size, etc. On the other hand, access to food and its utilization could be constrained by economic growth, lack of job opportunities, lack of credit, inadequate training, inadequate knowledge, etc., Hoddinott (1996). Different conversion factors were used to convert the available grain to total calories available for each household. The food supply at household level calculated in above step was used to calculate calories available per kilogram per person per day for each household. Using the FAO (2002), 2100 kilo calories per person per day was used as a measure of calories required (i.e. demand) to enable an adult to live a health and moderately active life.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

This chapter presents the mathematics tools that help in analysing data in the subsequent parts of the study. It contains four sections namely; sources of data, estimation of the food intake method, the specification of the regression model and specification of the probability model.

3.2 Study Area

The study site constitutes the Lake Victoria basin.

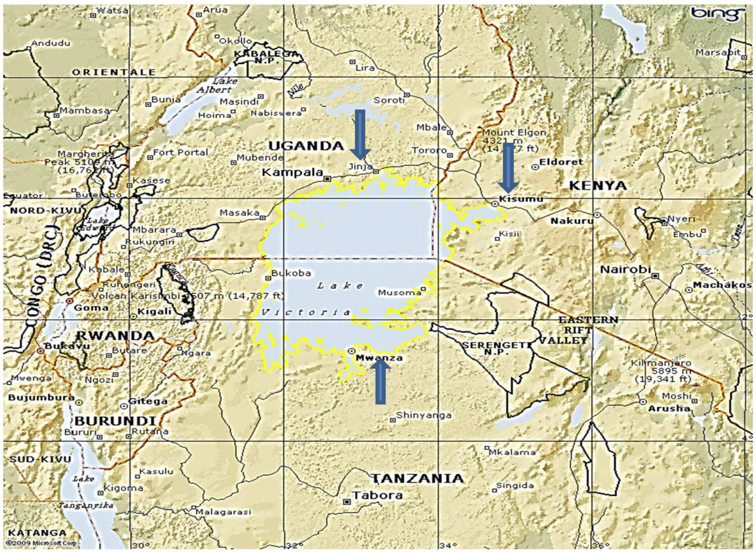

Figure 3.1: The Lake Victoria Basin

3.3 Sampling Strategy

The sampling frame was composed of clusters selected by (PPS) from a set of 100 households). At stage II the clusters sampled from each district were selected with equal probability. Therefore, the first stage was a de facto PPS sub-sample selection of a household. This sampling strategy produced an approximately self weighting sample of households in each stratum. From this a total sample of 135 households (45 households in each of 3 Primary Sampling Units). This sample design facilitated representative estimates at district, location and sub-location level, as well as in the third stage, which involved calculation of sampling selection probabilities of each selected household.

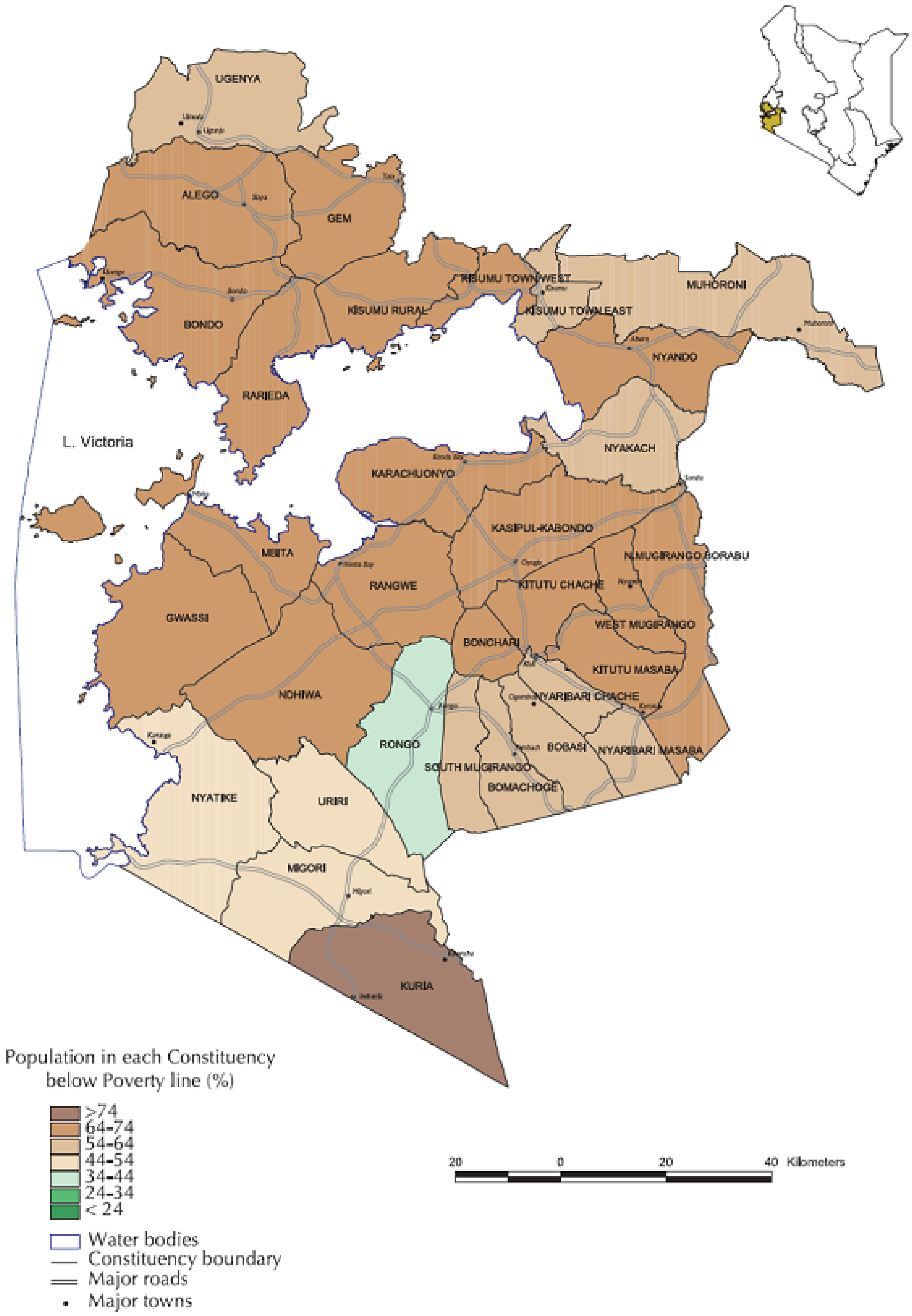

Figure 3.2: The Lake Victoria Basin on the Kenyan side

The probabilities were used to derive sampling weights needed to compute unbiased estimates and statistics presented in this study.

These lists were each used to select 15 households from each sub-location by employing simple random sampling technique. The sample units were chosen using proportionate sampling method. Household security depends on the factors such as food availability, socioeconomic condition of the society and procurement strategies, the present study uses variables, such as household demography, cereal production, food aid/assistance, source of income, household and animal assets, access to service, food consumption, nutritional status and dietary diversity collected during first year work of the project (Mwita *et al.*, 2007).

3.3.1 Data Source and Acquisition Methods

The study is based on both primary and secondary sources of information. Primary data was collected through survey, focus group discussions, and field observations. Secondary data were collected from governmental organizations both at regional and district level. The sources and methods used to acquire data for the research are outlined below.

(a) Primary data

Most of the data required to answer and validate the research questions were collected from primary sources. To generate the required data from the primary sources, different methodological approaches such as in-depth interviews, focus group discussions, and field observations were employed. These techniques were used to collect data pertaining household demographic characteristics, main livelihoods, asset ownership, income, crop production, household coping strategies, farm input use, food security status of households, opinions and understandings of households as to how the government policy intervention has addressed to the problem.

1. **Interview**

   These were used at the ward and division levels in order to obtain the primary data being collected by the district, where individual vendor stockiest were interviewed also at district markets the wholesalers/buyers were interviewed to give their food balance sheet during closing and opening their stocks. About 80 interviewees were involved in this exercise at all levels of ward, division and district.

2. **Focus Group Discussions**

   This methodology was focusing to the wholesales in order to gather the information views in relation to the food balance sheet they practice in their business.

   Then focused groups made the discussion focusing on food balance sheet if they have any information concerning it. The groups were able to share their experiences and give some reliable statistical figures in relation to food importation versus customers demand throughout the year also they were able to give the means on how they practice storage of their commodities to all levels to maintain food security off season and on season of the year.

3. **Observation**

   Observations of the people's way of life, their assets and resources, the ups and downs to overcome their daily struggles, their activities for living, etc, would provide valuable and supportive information. Having a good look at the physical and socio-economic infrastructures, the different economic activities people are involved with and government intervention programs currently undertaken would provide valuable contributions to understand the existing real situations and the overall situation of the poor. Thus, in this study an attempt was made to carefully observe every situation and understand them fully. Besides, direct field observation was employed as one of the methods to look how the policy programs are integrated with environmental sustainability.

(b) Secondary Data

This is statistical information which was collected from all levels and processed under district crop officer who is responsible with marketing and importation of food crops. The compilation of the FBS requires basic data on production, stocks, foreign trade, domestic utilization, nutrient value, dietary allowances and population which are obtained from results of census, household and established surveys, administrative reports of government agencies and special studies by various research institutions. The other part of the data was obtained from the country population census carried out in 2009 by the Kenya National Bureau of Statistics (KNBS).

3.4 Food Energy Intake

Non-parametric methods shows the "shape" of the relationship between Y and X variables without any parameters.The simplest method of density estimation is to divide the range of X into smaller number of intervals and count the number of times X is observed in each interval: i.e, the histogram.

when describing the data with a histogram there is an "art" choosing the number of "bin" or column. There are two problems with histograms, these are:

1. For a given number of bins,moving their exact location (boundary point) can change the graph.

2. The density function produced is a step function and the derivative either equal to zero or it's not defined.

This is a big problem if we are trying to maximize a likelihood function that is defined in terms of the densities of the distribution. Now let's define histogram more formally. First, define the density function for a variable X ;

For a particular value of X , call it X_0, the density function is:

$$
\begin{aligned}
f(x_0) &= lim_{h \to 0} \frac{F(x_0+h) - F(x_0-h)}{2h} \\
&= lim_{h \to 0} \frac{Prob[x_0 - h < x < x_0 + h]}{2h}
\end{aligned}
\tag{3.1}
$$

For a sample of data on x of size N, a histogram with a column width of $2h$, centering the column around x_0 can be approximated by:

$$
\begin{aligned}
\hat{f}_{HIST}(x_0) &= \frac{1}{N} \sum_{i=1}^{N} \frac{I(x_0 - h < x_i < x_0 + h)}{2h} \\
&= \frac{1}{Nh} \sum_{i=1}^{N} \frac{1}{2} I\left(\left| \frac{x_i - x_0}{h} \right| < 1 \right)
\end{aligned}
\tag{3.2}
$$

where $I(.)$ is an indicator function that is equal to 1, if the expression is true and 0 if it is false. Intuitively, this function equals the fraction of the sample that lies within h of x_0, divided by the column with $(2h)$.

3.4.1 Kernel Density Estimation

The first problem with histogram, is the arbitrarities in the location of the bin cutoff points;can be avoided by having a "moving" bin that is defined for every possible value of x. This can be done by replacing x_0 with x in the formula for $\hat{f}_{HIST}(x_0)$. Intuitively, for any hypothetical point x, this expression "count" how many actual data points the $x_i^{\prime}s$ are within $\frac{h}{2}$ of the hypothesis points and " normalizes" this count by number of observation (n) and the "bandwidth"h.

Then there is the problem of discontinuities in the function. Kernel estimation avoids discontinuities in the estimated (empirical) density function. In the above histogram formula, the "kernel" is everything to the right of the summation sign. The general formula is :

$$\hat{f}(x_0) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x_i - x_0}{h}\right) \tag{3.3}$$

where K (the kernel) is some function and h is a smoothing parameter ("bandwidth").

3.4.1.1 Kernel Assumptions

To get a continuous density function that integrates to 1 the kernel function $K(u)$ must satisfy;

- Continuous and symmetric around zero
- $\int_{-\infty}^{\infty} k(u)\,du = 1$, $\int_{-\infty}^{\infty} uk(u)\,du = 0$ and $\int_{-\infty}^{\infty} |k(u)|\,du < \infty$
- $k(u) = 0$ if $|u| \geq u_0$ for some u_0 or $|u|\,k(u) \to 0$ as $|u| \to \infty$
- $\int_{-\infty}^{\infty} u^2 k(u)\,du = \tau$ for some constant τ

3.4.1.2 Statistical Inference

In general, for an independent and identically distributed sample of the variables, X, for any value x_0, $\hat{f}(x_0)$ is a biased estimate of $f(x_0)$. The bias goes to zero if $h \to 0$ as $N \to \infty$, so $\hat{f}(x_0)$ is consistent. The bias depends on h, the curvature of $f(x)$ and the kernel chosen according to the following formula;

$$\begin{aligned} bias\left(\hat{f}(x_0)\right) &\equiv E\left[\hat{f}(x_0 - f(x_0))\right] \\ &= \frac{1}{2}h^2 f''(x_0)\int_{-\infty}^{\infty} u^2 k(u)\,du \end{aligned} \tag{3.4}$$

This implies that the "size" of this bias is $O\left(h^2\right)$. Assuming that $h \to 0$ as $N \to \infty$ the variance of $\hat{f}(x_0)$ is

$$Var\left[\hat{f}(x_0)\right] = \left(\frac{1}{Nh}\right) f(x_0)\int_{-\infty}^{\infty} (k(u))^2\,du + 0\left(\frac{1}{Nh}\right) \tag{3.5}$$

Thus the variance depends on the sample size, the bandwith(h), the density and the kernel function. It will go to 0 at a slower rate than N goes to infinity. The kernel estimates $\hat{f}(x_0)$ is point-wise consistent at any point x_0. If both the variance and bias disappear as $n \to \infty$, which requires that $h \to 0$ and $Nh \to \infty$, the uniform convergence property which is stronger holds if ${}^{Nh}\!/_{log_e h \to \infty}$.

The kernel estimation under certain conditions $\hat{f}(x_0)$ is also asymptotically normally distributed (Cameron and Trivedi, 2013).

3.4.1.3 Choice of Bandwidth

In general,large bandwidth reduces the variance by smoothing over a large number of points but this is likely to lead to bias because the points are averaged in a mechanical way that does not account for the particular shape of the distribution. In contrast, small bandwidth gives higher variance but have less bias. To undertake the tradeoff balance between bias and minimizing the variance, in theory, the natural approach is to minimize the mean squared error (MSE), which is the sum of the variance and the square of the bias. For any given point, x_0, the mean-squared errors is:

$$MSE\left(\hat{f}(x_0)\right) = Var\left[f'(x_0)\right] + \left[bias\left(f'(x_0)\right)\right]^2$$

$$= E\left[\hat{f}(x_0) - E\left[\hat{f}(x_0)\right]^2\right] + \left(E\left[\hat{f}(x_0) - f(x_0)\right]\right)^2 \tag{3.6}$$

$$E\left[\left(\hat{f}(x_0)\right)^2\right] - \left(E\left[\hat{f}(x_0)\right]\right)^2 + \left(E\left[\hat{f}(x_0)\right]\right)^2 - 2f(x_0)E\left[\hat{f}(x_0)\right] + \left(f(x_0)\right)^2$$

$$= E\left[\left(f(\hat{x_0})\right)^2\right] - 2f(x_0)E\left[\hat{f}(x_0)\right] + \left(f(x_0)\right)^2$$

$$= E\left[\left(\hat{f}(x_0) - f(x_0)\right)^2\right] \tag{3.7}$$

as shown, the bias is $O\left(h^2\right)$ and the variance is $O\left(\frac{1}{Nh}\right)$. Intuitively ,$h$ should be chosen so that the square of the bias and the variance are of the same order.the square of the bias is $O\left(h^4\right)$, so this implies that h satisfy $h^4 = \frac{1}{Nh}$, which implies $h = \left(\frac{1}{N}\right)^{1/5}$.

This implies that $h = O\left(N^{-0.2}\right)$ and $\sqrt{Nh} = O\left(N^{0.4}\right)$.

The ideas is that, we want to minimize the sum of the squared errors at a very large number of hypothetical points. As the number of points goes to infinity, this amount minimizes the mean of the intergrated squared errors (MISE). That is, an optimal bandwidth minimizes:

$$\begin{aligned} MISE(h) &= E\left[\int_{-\infty}^{\infty} \left(\hat{f}(x_0) - f(x_0)\right)^2 dx_0\right] \\ &= \int_{-\infty}^{\infty} MSE\left[\hat{f}(x_0)\right] dx_0 \end{aligned} \tag{3.8}$$

Differentiating $MISE(h)$ with respect to h and setting the derivative equal to zero yields the optimal bandwidth.

$$h^* = \delta \left[\int_{-\infty}^{\infty} \left(f''(x_0) \right)^2 dx_0 \right]^{-0,2} N^{-0.2} \tag{3.9}$$

Where δ depends on the kernel function used. The result shows that optimal bandwidth decreases (very slowly) as the sample size (N) increases.

$$h^* = O\left(N^{-0.2}\right)$$

This implies that, if the true density function has a lot of curvature (f'' is large), the bandwidth should be smaller.

3.5 Specification of the Regression Model

When poverty is defined as the current consumption deficit, a household is categorized as poor if the value of per capita consumption of its members is lower than the poverty line. Therefore, it is logical to search for poverty predictors based on variables that correlate with per capita household consumption. These variables can be obtained by estimating a model of consumption correlates, where the left-hand side is per capita consumption and the right-hand side is a set of variables that is thought of correlating with household consumption. Different from determinants model, in correlates model the endogeneity of the right-hand side variables is not a concern (Maddala, 1983).

Once the set of the right-hand side variables has been determined, a stepwise regression procedure is employed to estimate the model. The stepwise estimation procedure is used because in the end we want to obtain a manageable number of variables that can be relatively easily collected in practice and at the same time meaningfully used to predict household consumption level and poverty status.

3.5.1 The Augmented model

Model c_j, the determinants of per capita consumption at the household level using the simplest form of a model as follows

$$logc_j = \beta_j x_j + e_j \tag{3.10}$$

where x_j is a set of household characteristic and e_j is a random error term.the consumption model above can be described as the basic model.it has the feature that the marginal effects of the determinants of consumption are constant across households. It is however arguable that there is heterogeneity across households and the marginal effects themselves depend on household characteristics. This concern leads us to consider the augmented model that allows for a range of interaction effects and individual specific marginal effects (β_j);

$$logc_j \quad = \quad \beta_j x_j + e_j \tag{3.11}$$

where $\beta_j = \beta' + x_j + \varepsilon_j$ and hence

$$logc_j = \beta'_j x_j + x_j \phi x_j + e^*_j \tag{3.12}$$

This delivers a model with heteroscedastic errors ,$e^*_j = e_j + \varepsilon_j$, which is easily allowed for estimating the variance matrix of the model parameters. The model has a generalized quadratic form which is a numerically equivalent second order approximation to any arbitrary twice differentiable function Hosmer and Lemeshow (2000).

In general Equation 3.12 can be written in the form

$$Y = f(x, \beta) + e \tag{3.13}$$

where $\beta = \beta_1, \beta_2, \ldots \beta_p$ in a vector of p unknown parameter and $f(x)$ in a $p \times 1$ vector whose first element is equal to one and its remaining $p-1$ elements are polynomial functions of $x_1, x_2, \ldots x_k$. These functions are in the form of cross production of the x_j.

For example, the model in Equation 3.13 can be written as :

$$Y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{i<j} \beta_{ij} x_i x_j + e^*_j \tag{3.14}$$

where $f(x) = 1, x_1, x_2 \ldots x_k, x_1 x_2, x_1 x_3, \ldots x_{k-1} x_k$

The model in vector and matrix form is

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{n,p-1} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} e^*_j = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

representing the form

$$\mathbf{Y} \ = \mathbf{X}\beta + \mathbf{e}^* \tag{3.15}$$

Where

Y is a $(n \times 1)$ vector of response

β is a $(p \times 1)$ vector of parameters

X is a $(n \times p)$ design matrix

e_j^* is a $(n \times 1)$ vector of the error term.

σ^2 is a $(n \times 1)$ random vector of variance - Covariance matrix of the response variable.

$E\left[e_j\right] = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ and $\sigma^2(e) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$

3.5.2 Estimation of Regression Coefficient

In order to estimate β in the model 3.13, a series of n experiments $(n > p)$ are carried out in each of which the response Y is observed at different settings of the control variables; X_1, X_2,..., X_k.

Let Y_u denote the observed response value at X_u, where $X_u = (x_{u1}, x_{u2}, \ldots, x_{uk})$ with X_{ui} denoting the u^{th} setting of X_i at the u^{th} experimental run $(i = 1, 2, \ldots, k; u = 1, 2, \ldots, n)$.

From Equation 3.13 we then have

$$Y_u = f(X_0, \beta) + e_u \tag{3.16}$$

$u = 1, 2, \ldots, n$

Where e_u is the experiment error associated with $Y_u, (u = 1, 2, \ldots, n)$.

Model 3.16 can be experiment in matrix form as ;

$$Y = X\beta + e \tag{3.17}$$

Where $Y = (Y_1, Y_2, \ldots Y_n), X$ is an $n \times p$ matrix where u^{th} row in $f(X_u)$ and $e = (e_1, e_2, \ldots, e_n)$. The matrix X is assumed to be of full column rank, that is, rank $(X) = p$. In this case, model 3.17 is said to be of full rank.

In addition, it is assumed that $E[e] = 0$ and $var(e) = \sigma^2 I_n$, where σ^2 is unknown and I_n is the identity matrix of order $n \times n$. This implies that the response values $y_1, y_2, \ldots, y_n$ are uncorrected and have variance equal to σ^2 . Thus, the expected value of Y in $E[Y] = X\beta$ and the variance-covariance matrix is $Var(Y) = \sigma^2 I_n$.

Under the above assumption, estimation of β in model 3.17 can be achieved by using the method of ordinary least square (OLS). By definition the OLS of β denoted by $\hat{\beta}$ is the vector that minimises the square of the Euclidean norm of $Y - X\beta$, that is ,

$$\begin{aligned} S(\beta) &= \|Y - X\beta\|^2 \\ &= (Y - X\beta)'(Y - X\beta) \\ &= Y'Y - 2\beta'X'Y + \beta'X'X\beta \end{aligned} \tag{3.18}$$

Since $S(\beta)$ has first order partial derivative with respect to the element of β, a necessary condition for $S(\beta)$ to have a minimum at $\beta = \hat{\beta}$ in that $\frac{\partial(S(\beta))}{\partial\beta} = 0$ at $\beta = \hat{\beta}$, that is ;

$$\left[\frac{\partial}{\partial\beta}\left(Y'Y - 2\beta'X'Y + \beta'X'X\beta\right)\right]_{\beta=\hat{\beta}} = 0 \tag{3.19}$$

Applying theorem and corollary in Maddala (1983),we can write that

$$\frac{\partial}{\partial\beta}\left(\beta'X'Y\right) = X'Y \tag{3.20}$$

$$\frac{\partial}{\beta}\left(\beta'X'X\beta\right) = 2X'X\beta \tag{3.21}$$

Making substitution in Equation 3.19, we obtain

$$-2X'Y + 2X'X\hat{\beta} = 0 \tag{3.22}$$

Solving Equation 3.22 for $\hat{\beta}$, after noting that $X'X$ in a singular matrix by the fact X is a full column rank, we get

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y \tag{3.23}$$

Note that Equation 3.22 achieves the absolute minimum over the parameter space of β at $\hat{\beta}$ since Equation 3.19 has a unique solution given by $\hat{\beta}$ and the Hessian matrix of second order partial derivatives of $S(\beta)$ with respect to the element of β, namely the matrix

$$\begin{aligned}\frac{\partial}{\partial\beta'}\left[\frac{\partial}{\partial\beta}[S(\beta)]\right] &= \frac{\partial}{\partial\beta'}\left[-2X'Y+2X'X\hat{\beta}\right] \\ &= 2X'X \end{aligned} \tag{3.24}$$

is positive definite.

3.5.3 Properties of Ordinary Least-Square Estimators

Consider model 3.17 under the assumption that $E[e]=0$ and $Var(e)=\sigma^2 I_n$. Thus, $E[Y]=X\beta$ and $Var(Y)=\sigma^2 I_n$.

A number of results and properties associated with β are discussed in this section.

1. $E\left[\hat{\beta}\right]=\beta$, that is ; $\hat{\beta}$ is an unbiased estimation of β (proof easily obtained).
2. $Var\left(\hat{\beta}\right)=\sigma^2(X'X)^{-1}$
3. $\hat{\beta}\sim N\left(\beta,\sigma^2(X'X)^{-1}\right)$

   Since $\hat{\beta}=(X'X)^{-1}X'Y$, which is a linear function of Y and Y is normally distributed as $N\left(X\mid s,\sigma^2 I_n\right)$, then using (a) and (b) we conclude

   that $\hat{\beta}\sim N\left(\beta,\sigma^2(X'X)^{-1}\right)$

3.6 Specification of the Poverty Logistic Model

Choosing an appropriate model and analytical technique depends on the type of variable under investigation. Methods and regression models deal with cases where the dependent variable of interest is a continuous variable which we assume, perhaps after an appropriate transformation, to be normally distributed. But in many applications, the dependent variable of interest is not on a continuous scale; it may have only two possible outcomes and therefore can be represented by an indicator variable taking on values 0 and 1.

In this study, the dependent variable Y was defined to have two possible outcomes:

1. The households is poor if the threshold is < 2100 kilo calories per day (1).
2. The households is not poor if the threshold is > 2100 kilo calories per day (0).

These two outcomes are coded 1 and 0 respectively. This shows that the dependent variable was dichotomous and it can be represented by a variable taking the value 1 with probability π and the value 0 with probability $1-\pi$. Such a variable is a point binomial variable, that is, a binomial variable with $n=1$ trial, and the model often used to express the probability π as a function of potential independent variables under investigation is the logistic regression model. Therefore, to sort out which explanatory variables are most closely related to the dependent variable, in this study nine factors are considered. The method used in this study involves a linear combination of the explanatory or independent variables. Thus, the study was modeled within the framework of theories mentioned in the sections 3.4 and 4 and the model used by this study to determine factors affecting poverty status is given below.

3.6.1 Logistic Regression

The function has been discussed by many researchers like Fan *et al.* (1998). It is given by;

$$\begin{aligned} f(g) &= \frac{exp(g)}{1+exp(g)} \\ &= \frac{1}{1+exp(g)} \end{aligned} \tag{3.25}$$

when modeling a Bernoulli random variable with multi-variables, one directly models the probabilities of group membership, as follows;

$$P(Y=1|X=x) = \frac{1}{1+exp\left(-\left(\beta_0+\sum_{j=1}^{d} X_{ij}\beta_j\right)\right)} \tag{3.26}$$

where g in 3.25 is given by

$$g(X;\beta) = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_d X_{1d} \tag{3.27}$$

To illustrate, the applicability of the logistic function, the bold curve in the figure 3.3 shows that the logistic function puts more weight on the tails than the normal distribution.

The logistic model is bounded between zero and one, this property estimates the possibility of getting estimated or predicted probabilities outside this range which would not make sense (Hosmer and Lemeshow, 2000).

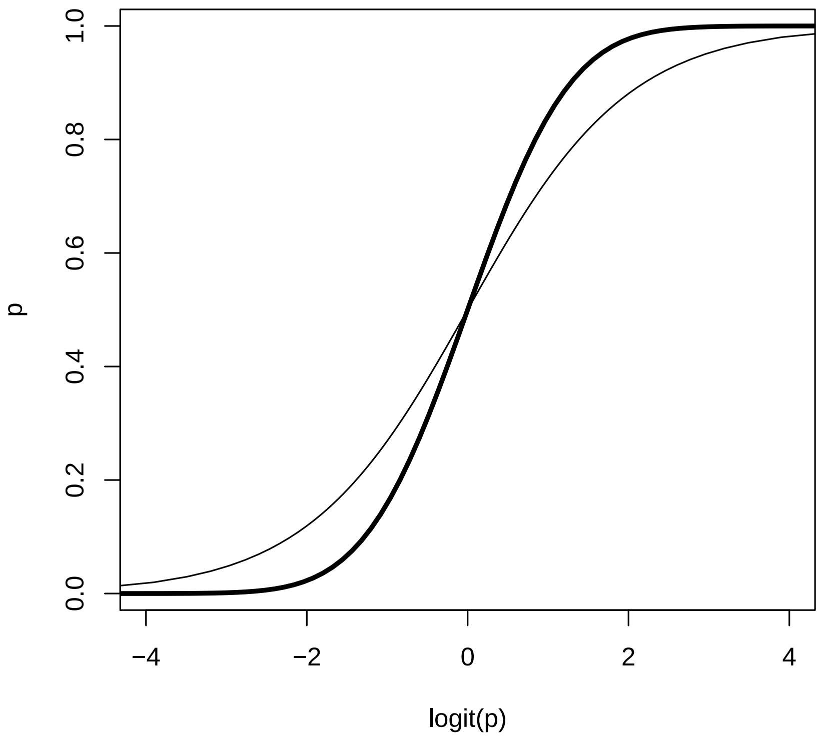

Figure 3.3: Standardized Normal and Logistic CDF's

Also with a proper transformation, one can get a linear model from the logistic function (Fahrmeir and Kaufmann, 1985). Fan *et al.* (1998) uses the logit function of the Bernoulli distributed response variable. Transforming 3.26 as in Fan *et al.* (1998) we have ;

$$\begin{aligned} Logit\left[P\left(Y=1|X=x\right)\right] &= log_e \frac{P\left(Y=1|X=x\right)}{1-P\left(Y=1|X=x\right)} \\ &= log_e \left\{ \frac{1+exp\left(\beta_0+\sum_{j=1}^{d}\beta_j X_{ij}\right)}{1+exp\left(-\left(\beta_0+\sum_{j=1}^{d}\beta_j X_{ij}\right)\right)} \right\} \end{aligned} \tag{3.28}$$

$$
\begin{aligned}
&= log_e\left(exp\left(\beta_0+\sum_{j=1}^{d}\beta_j X_{ij}\right)\right) \\
&= \beta_0+\sum_{j=1}^{d}\beta_j X_{ij} \qquad (3.29)
\end{aligned}
$$

the function 3.29 is a generalized linear model (GLM) with d independent variables.

The motivation to the use of logistic model was that it follows the properties of the GLM. Lets define the hypothetical population proportion of cells for which $Y = 1$ as $\pi = P(Y = 1|X = x)$. Then the theoretical proportion of cells for which $Y = 0$ is $1-\pi = P(Y = 0|X = x)$. We estimate π by the sample proportions of cells for which $Y = 1$. In the GLM context, it is assumed that there exists a set of predictor variables, $X_{11}, X_{12}, \cdots, X_{1d}$, that are related to Y and therefore provides additional information for estimating Y. For mathematical reasons of additivity and multiplicity, logistic model is based on linear model for the log odds in favour of $Y = 1$.

$$
log_e\frac{\pi_i}{1-\pi_i} = \alpha + \sum_{j=1}^{d}\beta_j X_{ij} \qquad (3.30)
$$

thus

$$
\pi_i = \sum_{j=0}^{d}\beta_j X_{ij}
$$

where $\beta \in \Re^d$ of unknown parameters.

The logistic regression (logit link)

$$
\begin{aligned}
g(\pi_i) &= log_e\frac{\pi_i}{1-\pi_i} \qquad (3.31)\\
&= logit(\pi_i)
\end{aligned}
$$

and

$$
g^{-1}(g(\pi_i)) = \pi_i \qquad (3.32)
$$

thus the inverse of the logit function in terms of $(X;\beta)$ is given by;

$$\begin{aligned} g^{-1}(X;\beta) &= \pi_i \\ &= \frac{1}{1+exp\left(-\left(\beta_0+\sum_{j=1}^{d}x_{ij}\beta_j\right)\right)} \end{aligned} \tag{3.33}$$

This model can be rewritten as

$$logit(\pi_i) = \sum_{j=0}^{d}\beta_j X_{ij} \tag{3.34}$$

3.6.2 Testing Coefficients of Independent Variables

3.6.2.1 Wald Test Statistic

Wald test statistic for a given coefficient β was obtained by first computing a $z-$statistic as

$$z = \frac{\hat{\beta}}{Se} \tag{3.35}$$

where was the sample estimate of β and Se was the standard error of the estimate.

The $z-$statistic was then squared giving a Wald statistic which has a χ^2 distribution. When the p value of the Wald statistic was less or equal to 0.05, the variable was considered significant and was retained in the model. If the p value was higher than 0.05, it implied the variable did not make a significant contribution to the model and was therefore excluded from the model. The Wald test was more reliable when samples were sufficiently large. It was used in this study as there was no danger of bias given the big sample of 135 households used.

3.6.2.2 The Likelihood-Ratio Test

The likelihood-ratio test yields more reliable estimates than the Wald test when samples are small. It is obtained as the ratio of the maximized value of the full model likelihood function L_1 to the maximized value of the simple model likelihood function L_0

The likelihood ratio-test statistics

$$\begin{aligned} -2log\left(\frac{L_0}{L_1}\right) &= -2\left[Log(L_0) - Log(L_1)\right] \\ &= -2\left[L_0 - L_1\right] \end{aligned} \tag{3.36}$$

This log transformation yields a χ^2statistic which is preferred in backward stepwise regression analysis. When the p value of the likelihood-ratio test statistic was less or equal to 0.05, the variable was considered significant and was retained in the model. If the p value was higher than 0.05, it implied the variable did not make a significant contribution to the model and was therefore excluded from the model.

3.6.3 Goodness-of-Fit test

The Hosmer-Lemshow test statistic was used to assess the goodness of fit of the model to the data. It is a χ^2statistic which evaluates goodness of fit by first dividing the subjects into 10 ordered groups based on their probabilities. Those with probabilities less than 0.1 form group 1, those with more than 0.1 and less than 0.2 form group 2 and so on until those with 0.9 to 1.0 form group 10. Then a comparison is made of the observed value obtained in each group with the value as predicted by the logistic regression model. The desired goal is to show that the observed and predicted do not differ significantly. If the model is good then the subjects with success are classified in the higher deciles of risk while those with failure are classified in the lower deciles of risk (Hosmer and Lemeshow, 2000; Agresti, 2002). The desired outcome here is non-significance showing that the model prediction does not significantly differ from the observed. The null hypothesis here was that there is no difference between the predicted values using the model and the actual values of the dependent variable. If the p value of the Hosmer and Lemeshow goodness-of-fit statistic was less or equal to 0.05, we reject the null hypothesis. If it was greater than 0.05, we failed to reject the null hypothesis and concluded that the model estimates did fit the data well and explained much of the variance in the dependent variable. The higher the value of the test statistic the better the model fit.

CHAPTER FOUR

MODEL ESTIMATION

4.1 Introduction

This chapter presents parameter estimation of the logistic model and simulation studies on the properties of the estimators. It contains two sections namely; parameter estimation and simulation study.

4.2 Parameter Estimation

Fan *et al.* (1998) pointed out that estimating the function $P(Y=1|X=x)$ in Equation 3.26 is equivalent to estimating the function $g(X;\beta)$ in Equation 3.27. Parametric estimation of $g(X;\beta)$ can be found in Joanes (1994); Pastor-Barriuso *et al.* (1998, 2003) among other authors, they used the maximum likelihood estimation method. As they pointed out, one first defines the likelihood function. For the Bernoulli distribution case we have

$$L(Y,X;\beta)=\prod_{i=1}^{n}[P(Y=1|X=x)]^{y_i}[1-P(Y=1|X=x)]^{1-y_i} \tag{4.1}$$

So, taking the logarithm and upon simplification we have

$$l(Y,X;\beta)=\sum\{(Y_i-g(X;\beta)-log_e(1+exp(g(X;\beta)))\} \tag{4.2}$$

The regularity conditions requires that the MLEs of β satisfies the usual consistency and asymptotic normality properties (Amemiya, 1985; Gourienx and Monfort, 1981).

The optimization of the function in 4.2 with respect to the unknown vector β requires iterative techniques since first derivative is nonlinear in $\hat{\beta}$ and has no simple analytical solution for $\hat{\beta}$ (Maddala, 1983).

$$\begin{aligned} l'(Y,X;\beta) &= \sum_{i=1}^{d} y_i x_{ij} - n_i \frac{exp\left(\sum_{i=1}^{d}\beta_j X_{ij}\right)}{1+exp\left(\sum_{i=1}^{d}\beta_j X_{ij}\right)} x_{ij} && (4.3)\\ &= \sum_{i=1}^{d} y_i x_{ij} - n_i \pi_i x_{ij} && (4.4)\end{aligned}$$

In matrix form, 4.4 can be rewritten in the form;

$$l'(Y,X;\beta) = \sum_{i=1}^{n} (y_i - \pi_i)\mathbf{X} \tag{4.5}$$

The equation,

$$\pi_i = \frac{exp\left(\sum_{i=1}^{d} \beta_j X_{ij}\right)}{1 + exp\left(\sum_{i=1}^{d} \beta_j X_{ij}\right)} \tag{4.6}$$

is strictly increasing function (monotone) of β_j and approaches 0 as$\beta_j \to \infty$ and approaches n as $\beta_j \to \infty$. The second derivative of 4.3 is strictly negative for all $\beta_j's$ and as such the solution is a maximum (Beer, 2001; Shifa, 2009).

4.2.1 Newton-Raphson Algorithm

The Newton-Raphson method requires that the starting values be sufficiently close to the solution to ensure convergence. Under this condition the Newton-Raphson iteration converge quadratically to at least a local optimum. When the method is used to the problem of maximizing the likelihood function, it produces a sequence of values $\theta^{(0)}, \theta^{(1)}, \cdots, \theta^{(\theta)}$ that under ideal conditions converge to the MLEs $\hat{\theta}_{mle}$.

The motivation to the use of the method is that this approximation is valid provided the unknown parameter β^{j+1} is in the neighbourhood of β^{j}. Since $l(Y,X;\beta)$ corresponds to the objective function to be minimized, $l'(Y,X;\beta)$ represents the gradient vector, the vector of first order partial derivative and $J(\theta)$ to the negative of the Hessian matrix $H(\theta)$ which is a matrix of the second order derivative of the objective function $L''(Y,X;\beta)$. Then the Hessian matrix is used to determine whether the minimum of the objective function $l(Y,X;\beta)$ is achieved by the solution $\hat{\theta}$ for the equation $l'(Y,X;\beta) = 0$, that is, whether $\hat{\theta}$ is a stationary point of $l(Y,X;\beta)$. If this is the case the $\hat{\theta}$ is the maximum likelihood estimate of the matrix of θ the iterative formula for finding a maximum or minimum of a function $f(x)$ is given by ;

$$X^{(j+1)} = X^{(j)} - H_i^{-1} l'(\theta) \tag{4.7}$$

where H_i is the Hessian $f''\left(X_i^j\right)$ and $l'(\theta)$is the gradient vector, $f'(x)$ of $f(x)$ at the i^{th} iteration.

Then the i^{th} iteration is given by;

$$\hat{\theta}^{(j+1)} = \hat{\theta}^{(j)} - \left(H\left(\hat{\theta}\right)^{j}\right)^{-1} l'(\theta) \tag{4.8}$$

In other words,

$$\hat{\theta}^{(j+1)} = \hat{\theta}^{(j)} - \frac{l'(\theta)}{l''(\theta)} \tag{4.9}$$

which is the iterative generator.

But from 4.8

$$l'(Y,X;\beta) = \sum_{i=1}^{d} y_i x_{ij} - n_i \pi_i x_{ij} \tag{4.10}$$

In matrix form;

$$l'(Y,X;\beta) = \sum_{i=1}^{n} (y_i - \pi_i)\mathbf{X} \tag{4.11}$$

and the negative of the second derivative;

$$\begin{aligned} J(\beta) &= \frac{\partial^2 L(Y,X;\beta)}{\partial\beta\partial\beta'} \\ &= \sum_{i=1}^{n} \pi_i (1-\pi_i)\mathbf{X}'\mathbf{X} \end{aligned} \tag{4.12}$$

The matrix $J(\beta)$ is a $(p+1)\times(p+1)$ matrix. The matrix plays a key role in the estimation procedure and yields the logit estimates obtained by inverting the Hessian (or expected Hessian) matrix or the information matrix. Then the Newton-Raphson iterative solution of the system of equations can be used to obtain the solution of $\beta's$. At the i^{th} iteration, estimates are obtained as;

$$\hat{\beta}^{(i+1)} = \hat{\beta}^{(i)} - \left[J\left(\hat{\beta}^{i}\right)\right]^{-1} l'\left(Y,X;\hat{\beta}^{i}\right) \tag{4.13}$$

where the least square estimates of the $\beta's$ are used as initial estimates.

Continue applying Equation 4.13 until there is essentially no change between the elements of β from one iteration to the next. At that point, the maximum likelihood estimates are said to converge.

4.2.2 Challenges of Newton-Raphson

- If θ_0 is chosen sufficiently near $\hat{\theta}$, convergence is very fast.

- Another problem with the method is its' lack if stability.

This can be solved by using the method of Fishers Scoring which simply replaces the observed second derivative with its' expectation to yield the iteration.

$\hat{\beta}^{(i+1)} = \hat{\beta}^{(i)} - \left[I\left(\hat{\beta}^{i}\right)\right]^{-1} l'\left(Y,X;\hat{\beta}^{i}\right)$

In many cases $I\left(\hat{\beta}^{i}\right)$ is easier to calculate and $I\left(\hat{\beta}^{i}\right)$ is always positive. This generally stabilizes.

4.3 A Simulation Study

In this section we describe how a simulation study was setup to assess consistency and normality of the parameters.

4.3.1 Consistency of the ML Estimators

Nonlinear system of equations arise commonly in statistic. In some cases, there will be a naturally associated scalar function of parameters which can be optimized to obtain parameter estimates. The MLE cannot be written in closed form expression, thus substantially complicating the task of evaluating the characteristic of its (finite sample) distribution, whether the variables are random or not. Maximum likelihood estimator simulation for large samples are carried out using the Monte-Carlo simulation method. The simulations of the study involves the regressor variables which are fixed and for each model parameter, n-simulation binomial data set are generated for each of the regressor variable $x_1, x_2, \cdots, x_n$. We consider the complete model to be simulated as;

$$\begin{aligned} y_i &= g(X;\beta) + e_i \\ &= 1\, if\; X_i\beta + e \geq a \\ &= 0\, if\; X_i\beta + e < a \end{aligned} \tag{4.14}$$

where y_i is the dependent variable to incorporate the effects of the independent variables. The row vector X_i represents the i^{th} observations on all predictor variables.

The basic model can be structured as

$$\begin{aligned} \pi_i &= Pr(y_i = 1|x_i) \\ 1 - \pi_i &= Pr(y_i = 0|x_i) \end{aligned} \tag{4.15}$$

For the logit model;

$$\pi_i = \frac{exp\,(x'\beta)}{1 + exp\,(x'\beta)} \tag{4.16}$$

which is the cdf of the logistic distribution.

Consistency of the maximum likelihood estimator was assessed through standard Monte Carlo simulation, the finite sample of performance of consistency of the maximum likelihood estimators of the logistic regression model. In simulation study, four explanatory variables X_1, X_2, X_3 and X_4 which are fixed and the binary response variable Y,which is treated as a random variable in the logistic model were considered. For the fixed values of the intercept parameter β_0 and four other parameter $\beta_1, \beta_2, \beta_3$ and β_4. The aim was to compare the performance of the values of the parameters and their standard errors when the sample size increases. For fixed values of $\beta_0 = -22$, $\beta_1 = 2.5$, $\beta_2 = 0.15$, $\beta_3 = 0.03$ and $\beta_4 = 0.8$

$$\pi(x) = \frac{\exp(-22 + 2.5x_1 + 0.15x_2 + 0.03x_3 + 0.8x_4)}{1 + \exp(-22 + 2.5x_1 + 0.15x_2 + 0.03x_3 + 0.8x_4)} \tag{4.17}$$

In the simulation, sample sizes of $n = 200$, 300, 500 and 700 generate $5,000$ independent sets of random samples for each different sample size was considered. For each set of random samples with a particular sample size, we estimate $\beta_0, \beta_1, \beta_2, \beta_3,$ andβ_4 and their standard errors based on the logistic regression model. The final estimates and the standard errors of $\beta_0, \beta_1, \beta_2, \beta_3,$ and β_4 are the average of the 5,000 estimates of $\beta_0, \beta_1, \beta_2, \beta_3,$ and β_4 for that particular sample size. The following table gives the results of the simulation study for different sample sizes.

Table 4.1: Estimated parameter (β) values and their standard errors

β	$n = 200$		$n = 300$		$n = 500$		$n = 700$	
	Estimates	SE	Estimates	SE	Estimates	SE	Estimates	SE
β_0	-42.356	472.855	-24.268	12.830	-22.872	3.583	-22.497	2.947
β_1	6.237	73.156	3.425	2.173	3.236	0.436	3.177	0.357
β_2	0.310	4.085	0.136	0.079	0.129	0.057	0.127	0.047
β_3	0.039	0.716	0.036	0.024	0.034	0.018	0.033	0.015
β_4	1.501	2.204	0.983	0.29	0.937	0.206	0.920	0.169

As seen in the Table 4.1, for sample $n = 200$, the estimated values of the parameters are different from the true values ($\beta_0 = -22, \beta_1 = 2.5, \beta_2 = 0.15, \beta_3 = 0.03,$ and $\beta_4 = 0.8$)

and also the standard errors becomes larger. However, when the sample size increases from $n = 200$ to $n = 700$, the estimated values of the parameters are very close to the true values $\beta_0, \beta_1, \beta_2, \beta_3,$ and β_4, and the standard deviations of the estimates are noticeably smaller. This indicates that this simulation study performs well in showing the consistency of the maximum likelihood estimators for parameters of the logistic model.

4.3.2 Regularity conditions of the maximum likelihood estimates

In the 1920's, R.A Fisher originally developed the principle of maximum likelihood estimation (MLE) and established optimal properties of estimates by maximizing the likelihood function (Adrich, 1997).

The optimal properties in estimation are consistency (true parameter value that generated the data recovered asymptotically, that is, for data of sufficiently large samples); sufficiency (complete information about the parameter of interest contained in an MLE estimator): efficiency (lowest-possible variance of parameter estimates achieved asymptotically): and parameterization invariance (same MLE solution obtained independent of the parameterization used). Under certain regularity conditions, the MLE exhibits several characteristics that can be interpreted to mean that it is "asymptotically optimal" Lehmann and G.Casella (1998) provided the following results. In theorem FOUR.1 of the MLE regularity conditions. These conditions are;

- $A1$ The distribution P_θ of the observations are distinct (otherwise, θ can not be estimated consistently).
- $A2$ The distribution P_θ have a common support.
- $A3$ The random variables $X_i = (X_{i1}, X_{i2}, \cdots, X_{ip})$, $i = 1, 2, \cdots, n$, where the X_i are independent and identically distributed (iid) with probability density $f(X_i|\theta)$ with respect to probability measure μ.
- $A4$ There exists an open set ω of Ω containing the true parameter point θ° such that for almost all x, the density $f(X|\theta)$ admits all third derivatives

$$\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l} f(X|\theta) \qquad \text{for all} \quad \theta \in \omega \tag{4.18}$$

- $A5$ The first and second derivatives of $\log f$ satisfy the equations

$$E_\theta\left[\frac{\partial}{\partial\theta}\log f(X|\theta)\right] = 0 \quad \text{for} \quad j = 1, 2, \cdots, p \tag{4.19}$$

and

$$
\begin{aligned}
I_{ij} &= E_\theta\left[\frac{\partial}{\partial\theta_j}\log f(X|\theta)\cdot\frac{\partial}{\partial\theta_k}\log f(X|\theta)\right] \\
&= E_\theta\left[\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log f(X|\theta)\right]
\end{aligned}
\tag{4.20}
$$

- $A6$ Since the $p\times p$ matrix $I(\theta)$ is a covariance matrix, it is positive semi-definite. We will assume $I_{jk}(\theta)$ are finite and that the matrix $I(\theta)$ is positive definite for all θ into, and the p satisfies

$$
\frac{\partial}{\partial\theta_1}\log f(X|\theta),\cdots,\frac{\partial}{\partial\theta_p}\log f(X|\theta). \tag{4.21}
$$

are affinely independent with probability 1.

- $A7$ Finally, we will assume that there exist function M_{jki} such that

$$
\left[\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l}f(X|\theta)\right]\leq M_{jkl}(X)\qquad\text{for all}\quad\theta\in\omega \tag{4.22}
$$

where $M_{jkl}=E_\theta\left[M_{jkl}(x)\right]<\infty$ for all j, k, l.

Theorem FOUR.1. *Let $X_1,X_2,X_3,\cdots,X_n$ be iid each with a density $f(X|\theta)$ (with respect to μ which satisfies $(A1)-(A7)$) above. Then, with probability tending to 1 as $n\to\infty$, there exists solutions $\hat\theta_n=\hat\theta(X_1,\cdots,X_n)$ of the likelihood equations.*

$$
\frac{\partial}{\partial\theta_j}\left[f(X_1|\theta),\cdots f(X_n|\theta)\right]=0,\qquad j=1,2,\cdots,n \tag{4.23}
$$

or equivalently

$$
\begin{aligned}
\frac{\partial}{\partial\theta_j}\left[\log L(\theta)\right]=0,&\qquad j\\
=1,2,\cdots,p&
\end{aligned}
\tag{4.24}
$$

such that

(a) $\hat\theta_{jn}$ is consistent for estimating θ_j.

(b) $\sqrt{n}(\hat\theta_n-\theta)$ is asymptotically normal with mean (vector) zero and covariance matrix $[l(\theta)^{-1}]$, and

(c) $\hat{\theta}_{jn}$ is asymptotically efficient in the sense that

$$\sqrt{n}(\hat{\theta}_{jn} - \theta_j) \xrightarrow{L} N\left(0, [I(\theta)]_{jj}^{-1}\right) \tag{4.25}$$

4.3.3 Regularity conditions of the estimator in Logistic Regression

Fahrmeir and Kaufmann (1985) present regularity conditions for a very general class of generalized linear models. In this section, the regularity conditions under the Binomial response model were explained and then Theorem $FOUR.1$ was applied to show the asymptotic properties of ML estimators for the Binomial response model.

(C1): The pdf $g(X;\beta)$ is distinct, that is $\beta \neq \beta'$ implying that $g(X;\beta) \neq g(X;\beta')$, thus the model is identifiable.

The proof of this assumption has been well documented by Shifa (2009).

(C2): The pdf have common support for all β, the true parameter vector is in the interior of this space.

This condition holds if the domain (support) of X is a closed set (McFadden, 1974).

McFadden (1974) noted that the restriction that true parameter vector in the interior excludes some cases where consistent and asymptotically normal (CAN) breaks down. This is not a restrictive assumption in most application, but it is for some.

(C3): The response model is measurable in X, and for almost all X is continous in the parameters. The standard models such as the probit, logit and the linear probability model are all continous in their argument and in X, so that this assumption holds.

(C4): The model satisfies a global identification (that is it guarantees that there is at most one global maxima, see (McFadden, 1974).

The proof of this assumption has been discussed well by Shifa (2009). The concavity of the log-likelihood of an observation for the logit guarantees global identification, provided only that the $X's$ are not linearly independent.

(C5): The assumption states that the model log likelihood is twice or three times differentiable, this is true provided the parameters do not give observations on the boundary in the linear or log linear models where probabilities are zero or one. Deutsch (2007) shows that these conditions are specifically satisfied for the binomial model.

(C6): The log likelihood and its derivative have bounds independent of the parameters in some neighbourhood of the true parameter values. The first derivative have the Lipschitz property in the neighbourhood. This property is satisfied by the logistic model since it is continuously differentiable (McFadden,1999).

(C7): The pdf $g(X;\beta)$ is three times differentiable as a function of β . Further, for all $\beta \in \Omega$, there exists a constant c and a function $M(x)$ such that

for all $\beta_0 - c < \beta < \beta_0 + c$ and all x in the support of X.

$$\left| \frac{\partial^3}{\partial \beta^3} \log g(X;\beta) \right| \leq M(x) \tag{4.26}$$

with

$$E_{\beta_0}[M(X)] < \infty \tag{4.27}$$

for all $\beta_0 - c < \beta < \beta_0 + c$ and all x in the support of X. The proof of this assumption has been done by many authors like Beer (2001); Shifa (2009). This implies that the information matrix, equal to the expectation of the outer product of the score of an observation is non-singular at the true parameter.

The conditions $(C1), \cdots, (C7)$ may seem restrictive at first, but are met for a wide range of link functions. The results guarantee that the MLE estimates of β is essentially carried out by linearizing the first order condition for the estimator using a Taylor's expansion. Since the binomial model satisfies the above conditions, then following theorem holds for the parameter $\widehat{\beta}$.

Theorem FOUR.2. *Let $X_1, X_2, X_3, \cdots, X_n$ be iid each with a density $g(X;\beta)$. Then, with probability tending to 1 as $n \to \infty$, there exists solutions $\hat{\beta}_n = \hat{\beta}(X_1, \cdots, X_n)$ of the likelihood equations.*

$$\frac{\partial}{\partial \beta_j}[g(X_1;\beta), \cdots g(X_n;\beta)] = 0, \qquad j = 1, 2, \cdots, n \tag{4.28}$$

or equivalently

$$\frac{\partial}{\partial \beta_j}[\log L(\beta)] = 0, \qquad j = 1, 2, \cdots, d$$

such that

(a) $\hat{\beta}_{jn}$ is consistent for estimating β_j.

(b) $\sqrt{n}(\hat{\beta}_n - \beta)$ is asymptotically normal with mean (vector) zero and covariance matrix $[L(\beta)^{-1}]$, and

(c) $\hat{\beta}_{jn}$ is asymptotically efficient in the sense that

$$\sqrt{n}(\hat{\beta}_{jn} - \beta_j) \xrightarrow{L} N\left(0, [I(\beta)]^{-1}_{jj}\right)$$

4.3.4 Normality of the ML estimators

Under some assumptions that allows among several analytical properties, the use of the delta method, the central limit theorem holds. Simulation study was conducted through the freeware package R. The study shows how the properties of an estimator are affected by changing conditions such as its sample size and the value of the underlying parameters. Employing it in practice, The study illustrate the large sample behavior of the estimated parameters $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3,$ and $\hat{\beta}_4)'$ and also look at the sensitivity of the QQ-plots using the Shapiro-Wilks and the Kolmogorov-Smornov test, results show that;

$$\sqrt{N}\left(\hat{\beta_{mle}} - \beta\right) \longrightarrow N\left(0, \frac{1}{I(\beta_{mle})}\right) \tag{4.29}$$

where

$$I(\beta) = -E_\beta \begin{pmatrix} \frac{\partial^2 \log l}{\partial \beta_0^2} & \frac{\partial \log l}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 \log l}{\partial \beta_0 \partial \beta_1} & \frac{\partial \log l}{\partial \beta_0 \partial \beta_2} & \frac{\partial \log l}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 \log l}{\partial \beta_1 \partial \beta_0} & \frac{\partial \log l}{\partial \beta_1^2} & \frac{\partial^2 \log l}{\partial \beta_1 \partial \beta_2} & \frac{\partial \log l}{\partial \beta_1 \partial \beta_3} & \frac{\partial \log l}{\partial \beta_1 \partial \beta_4} \\ \frac{\partial^2 \log l}{\partial \beta_2 \partial \beta_0} & \frac{\partial \log l}{\partial \beta_2 \partial \beta_1} & \frac{\partial \log l}{\partial \beta_2^2} & \frac{\partial \log l}{\partial \beta_2 \partial \beta_3} & \frac{\partial \log l}{\partial \beta_2 \partial \beta_4} \\ \frac{\partial^2 \log l}{\partial \beta_3 \partial \beta_0} & \frac{\partial \log l}{\partial \beta_3 \partial \beta_1} & \frac{\partial \log l}{\partial \beta_3 \partial \beta_2} & \frac{\partial \log l}{\partial \beta_3^2} & \frac{\partial \log l}{\partial \beta_3 \partial \beta_4} \\ \frac{\partial^2 \log l}{\partial \beta_4 \partial \beta_0} & \frac{\partial^2 \log l}{\partial \beta_4 \partial \beta_1} & \frac{\partial^2 \log l}{\partial \beta_4 \partial \beta_2} & \frac{\partial \log l}{\partial \beta_4 \partial \beta_3} & \frac{\partial \log l}{\partial \beta_4^2} \end{pmatrix} \tag{4.30}$$

A quantile-quantile normal graph, plots the quantiles of the data set against the theoretical quantiles of the standard normal distribution. If the data set appears to be a sample from a normal population, then the points will fall roughly along the line. The computation results indicates that the distribution of parameters approximates normal distribution as sample size, n increases.

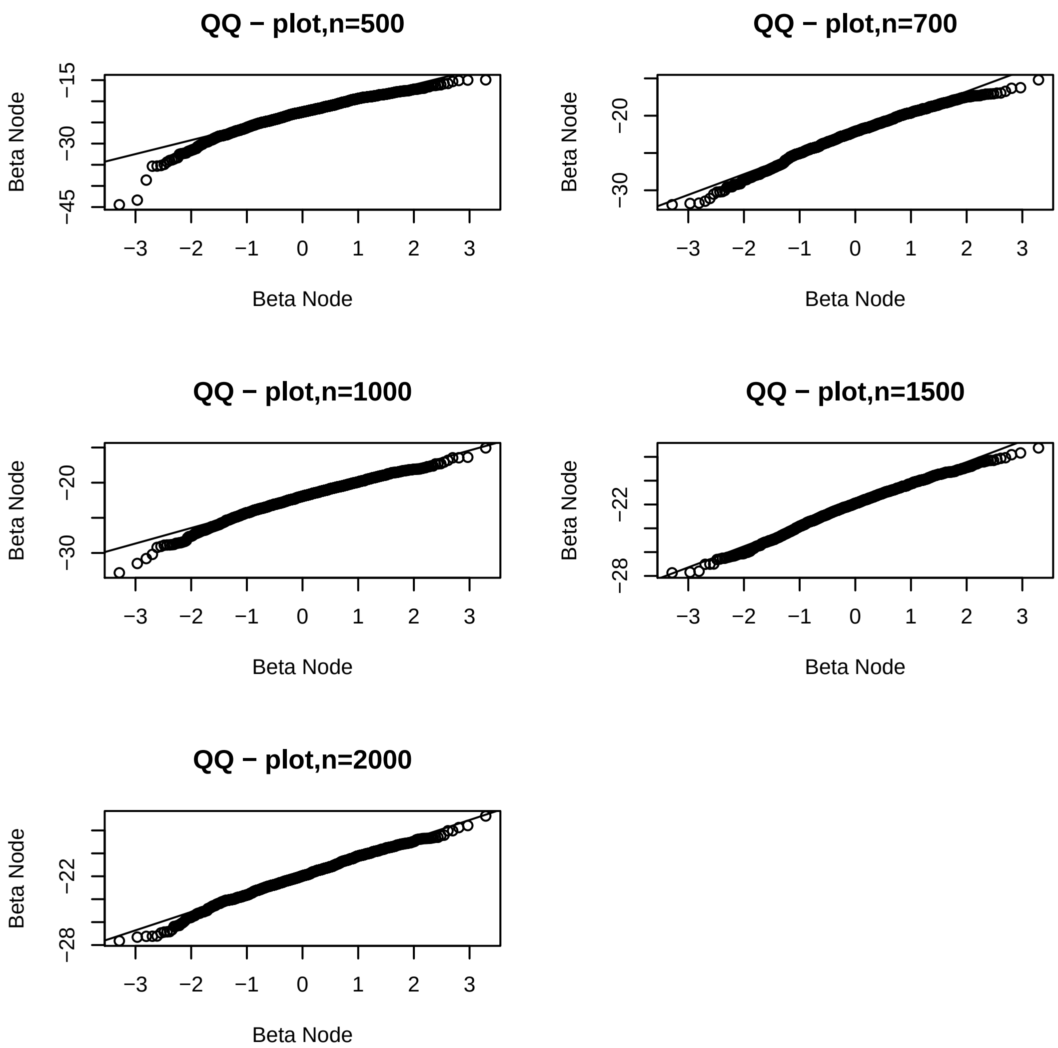

Figure 4.1: Monte Carlo Simulation of finite sample behaviour for normality of $\hat{\beta}_0$

Table 4.2: Test for Nomality, β_0

	Kolmogorov-Smirnov test		Shapiro-Wilks test	
Sample size(n)	Test statistic (D)	P-value	Test statistic (D)	P-value
500	0.0501	0.09864	0.9948	0.0016
700	0.0664	0.0100	0.9961	0.0119
1000	0.0389	0.3246	0.9964	0.01997
1500	0.0325	0.5600	0.9987	0.3417
2000	0.0323	0.5567	0.9986	0.0462

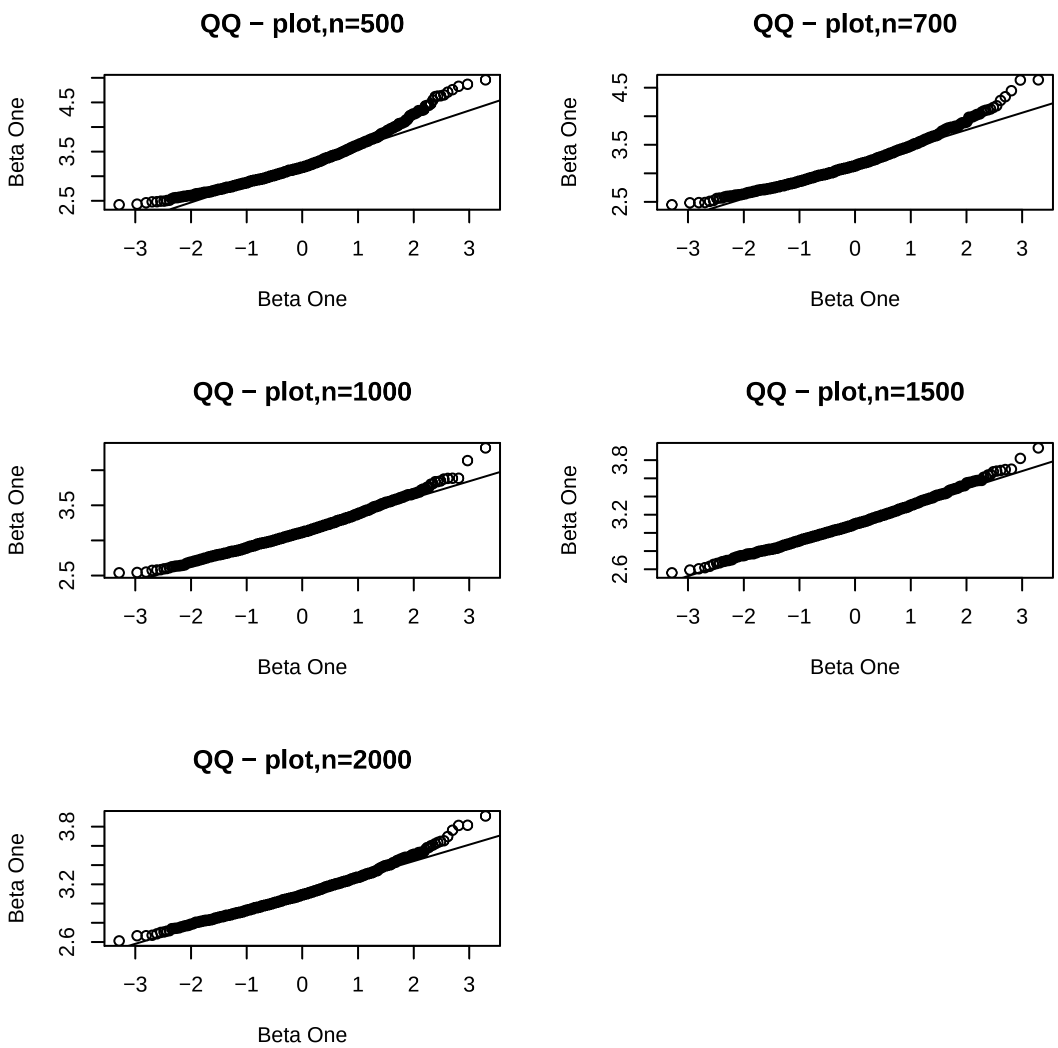

Figure 4.2: Monte Carlo Simulation of finite sample behaviour for normality of $\hat{\beta}_1$

Table 4.3: Test for Nomality, β_1

	Kolmogorov-Smirnov test		Shapiro-Wilks test	
Sample size(n)	Test statistic (D)	P-value	Test statistic (D)	P-value
500	0.0600	0.0015	0.0600	0.0015
700	0.0584	0.0037	0.0654	0.0004
1000	0.0493	0.0156	0.0493	0.0156
1500	0.0431	0.0491	0.0431	0.0491
2000	0.0312	0.2846	0.0312	0.2846

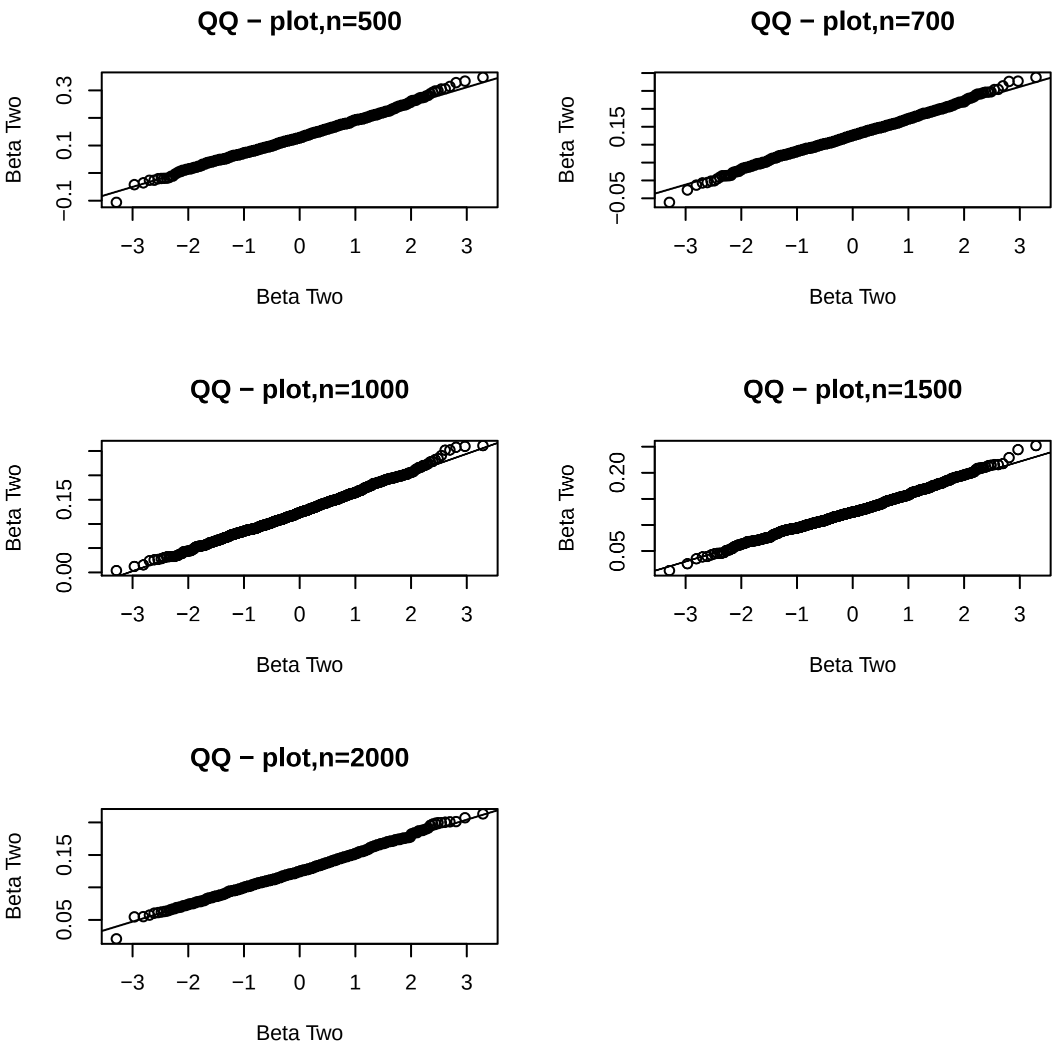

Figure 4.3: Monte Carlo Simulation of finite sample behaviour for normality of $\hat{\beta}_2$

Table 4.4: Test for Nomality, β_2

	Kolmogorov-Smirnov test		Shapiro-Wilks test	
Sample size(n)	Test statistic (D)	P-value	Test statistic (D)	P-value
500	0.0309	0.2948	0.9998	0.0899
700	0.0346	0.1831	0.9970	0.1934
1000	0.0326	0.2378	0.9961	0.05678
1500	0.0295	0.3457	0.9974	0.0403
2000	0.0291	0.3661	0.9995	0.1101

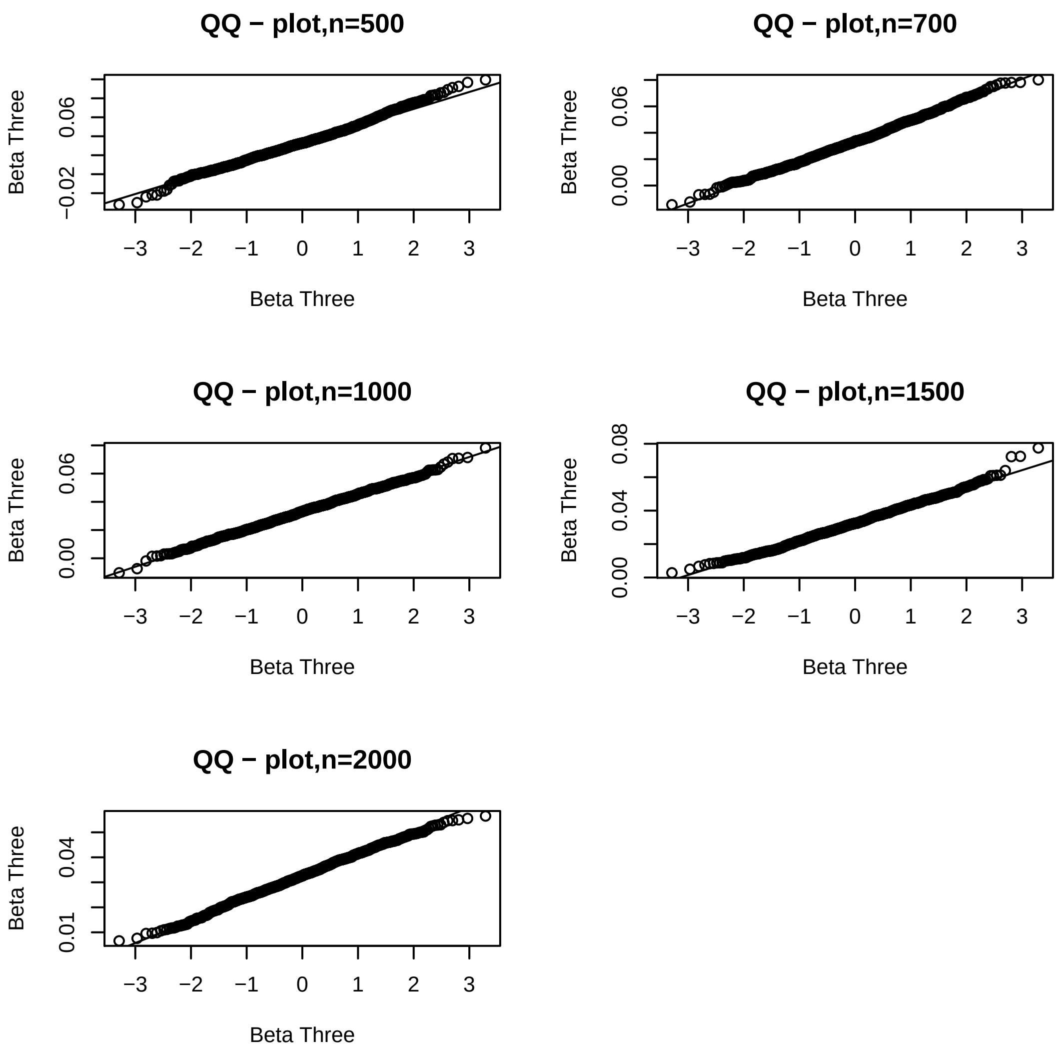

Figure 4.4: Monte Carlo Simulation of finite sample behaviour for normality of $\hat{\beta}_3$

Table 4.5: Test for Nomality, β_3

	Kolmogorov-Smirnov test		Shapiro-Wilks test	
Sample size(n)	Test statistic (D)	P-value	Test statistic (D)	P-value
500	0.0315	0.2731	0.9930	0.0001
700	0.0287	0.3825	0.9952	0.0029
1000	0.0167	0.5498	0.9969	0.0471
1500	0.0122	0.8700	0.9945	0.0707
2000	0.0096	0.8374	0.9988	0.7674

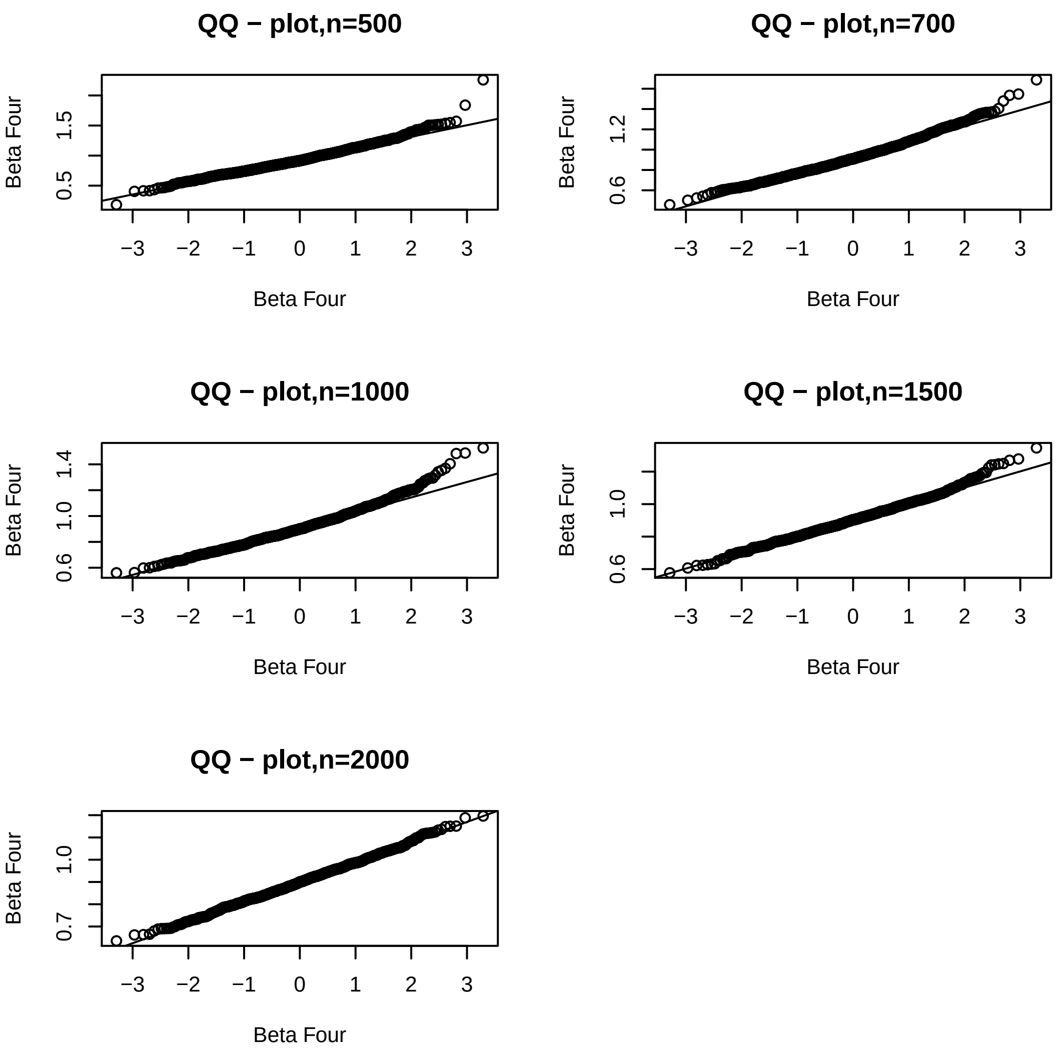

Figure 4.5: Monte Carlo Simulation of finite sample behaviour for normality of $\hat{\beta}_4$

Table 4.6: Test for Nomality, β_4

	Kolmogorov-Smirnov test		Shapiro-Wilks test	
Sample size(n)	Test statistic (D)	P-value	Test statistic (D)	P-value
500	0.0426	0.0529	0.9958	0.0084
700	0.0363	0.1431	0.9916	0.01843
1000	0.0459	0.2952	0.9968	0.04791
1500	0.0225	0.6946	0.9973	0.09807
2000	0.0187	0.9001	0.9995	0.9980

CHAPTER FIVE

DATA ANALYSIS

5.1 Empirical Studies on the Food Energy Intake Method

In this study, since the price data were not available, the method of food energy intake was used. The goal was to find the threshold level of consumption expenditure (or income) that allows the household to obtain enough food to meet its energy requirements.

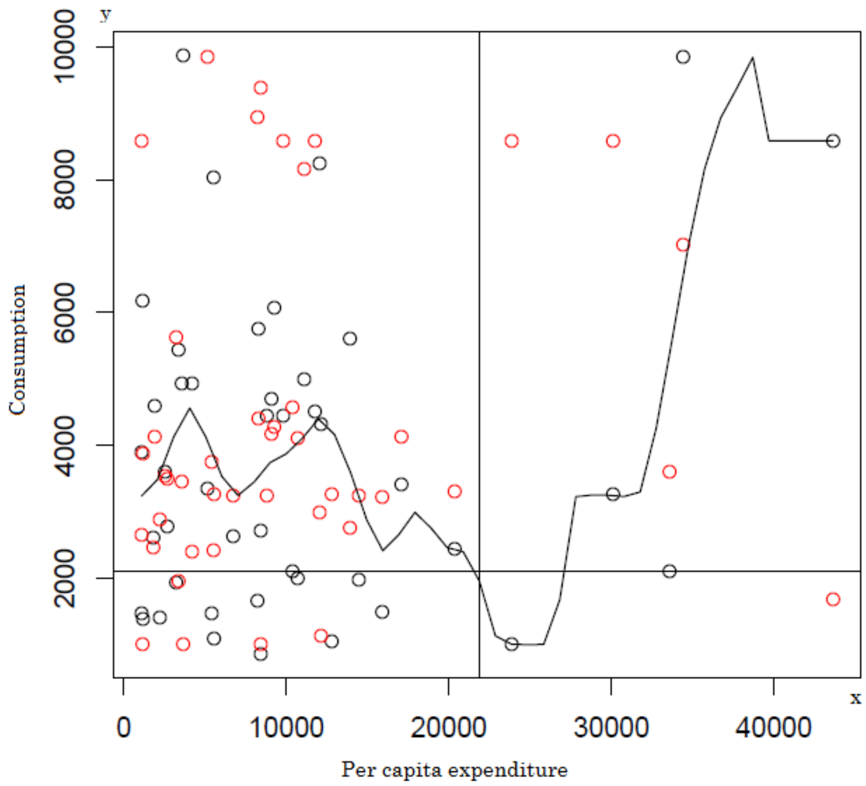

Figure 5.1: Calorie Income Function

Figure 5.1 shows the calorie income function as obtained in this study. It was obtained that as the income increases (or expenditure rises) food energy intake also rises although typically more slowly.

The key finding of this study are that the poverty income level was Ksh. 21,900 per month. In US dollar this was on average per person per day on a household of \$1.21, assuming an average of seven people in a household. The assumption of economies holds. This value was quite consistent with the United Nations threshold for many sub-Saharan Africa countries which is \$1.25 per person per day in a household (Ravallion and Bidani,

1994) . Other Poverty measures in the region that have also been used in the study are given in the Table 5.1.

Table 5.1: Poverty measures

Poverty measures	Percentage level (%)
Headcount index	31.250
Poverty gap index	9.325
Poverty severity index	2.696

From table 5.1 we see that 31.25% of the households in the region are poor, but as a welfare indicator the head-count index is unsatisfactory in that it violates the transfer principle (Ravallion, 1996b).

Dalton (1920), states that the transfer from a rich to a poor person should improve the measure of welfare. The head-count index provides a poor poverty estimate since estimates should be calculated for individuals, not households. The poverty gap index from the 5.1 indicates that 9.325% of the household fall below the poverty line. The poverty gap index also violates the transfer principle, thus not a good poverty measure.

The square poverty gap index measure explicitly puts more weight on the observations that fall below the poverty line, but does not give information on how poor the individuals of the household are. Thus, as seen in the three methods, the food energy intake method of obtaining the threshold in terms of expenditure and consumption is best (Ravallion, 2008).

5.1.1 The Gini Coefficient: A Measure of Inequality

In this study inequality refers to the dispersion of the distribution over the entire consumption aggregate.The widely used measure of inequality was the Gini Coefficient which ranges from zero (indicating perfect equality that is, where everyone in the population has the same expenditure or income) to one (indicating perfect inequality that is, when all expenditure or income is accounted for by a single person in the population). For most developing countries, the Gini Coefficient ranges between zero point three (0.3) and zero point six (0.6) (WB, 2006).

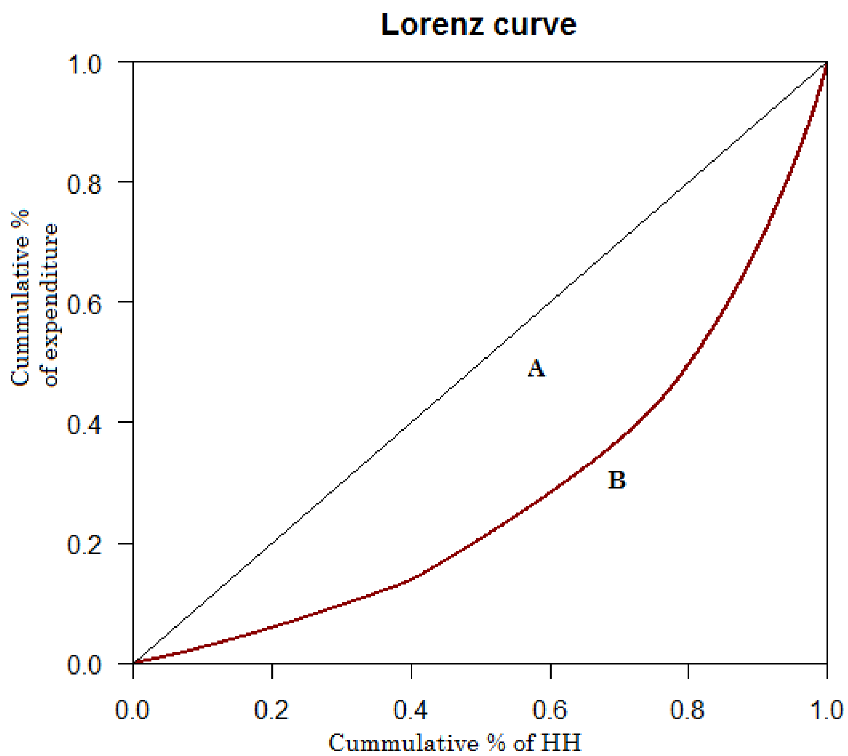

Figure 5.2: Lorenz Curve and Gini Coefficient

The Gini Coefficient of the region as obtained fromKNBS (2007c) the data is zero point four three two five (0.4325) which is very close to the value obtained using per adult equivalent for both 1997 welfare monitoring survey and KIHBS (2005/06) where in both this study and the two studies, the expenditure was used as a proxy to income.The rural Kenya Gini Coefficient of expenditure per adult equivalent was about point four one seven (0.417) during the same time (KNBS, 2007a).

5.2 Empirical studies of the Stepwise and the Augmented Regression Models

In this study several of the explanatory variables, there are observations with missing data and have constructed dummy variable that take a value of one if the household is missing data for a particular variable(while the value of that variable itself was set as zero). In this way potential of sample selection bias was reduced, without losing useful information from household with some valid data for most variables.

Per capita consumption was used as the basic measure of individual welfare.The use of per capita consumption imposes the assumptions that there are no economies of household size in consumption and that household composition does not matter.Therefore , the estimated parameters must be interpreted with caution.

There was also some concern of potential bias in parameter estimates due to endogeneity of omitted variables. If these factors are significant determinants of welfare ,the error term will not converge to zero in probability, and the parameter estimated for the individual explanatory variables will be inconsistent. To control this ,interactions term effects are included in the model.

While the augmented equation 3.12 offers a fairly general approach to modeling welfare, this generality comes with the potential cost of over-parameterizing the model with the full set of interaction terms, there are an explosion of parameters. Beginning with a k-parameters in the basic model, there are $\frac{2k+k(k-1)}{2}$ parameters in the augmented equation 3.12.

A model with numerous parameters was likely to suffer from multicollinearity. In the view of these difficulties; In this study, stepwise regression as a basic model was to limit these difficulties to only those significant variables in the model.

Table 5.2: Stepwise and augmented modeling of the log per capita

Variables	Description	Stepwise model		Augmented model	
		Coefficient	t-ratio	Coefficient	t-ratio
X1	HHsize	0.4079(.)	1.963	2.297(***)	5.466
X2	HHsize^2	-0.028(*)	-2.062	-0.2410(**)	-3.356
X3	Gender HH(head)	0.4988(.)	1.853		
X5	Landsize (acre)	0.5824(.)	1.983	0.5335(*)	2.203
X6	HH(head) age	0,1588(***)	3.575		
X7	HH(head) age^2	-0.0016(**)	-3.575		
X8	HH Average in school	0.0857(.)	1.868	0.2568(**)	2.886
X9	Production(kg) per year	0.0005(***)	1.890	0.0029(*)	2.676
X1:X2	HHsize* HHsize^2			0.0089(**)	2.956
X1:X8	HHsize* HH Average in school			-0.0277(*)	-2.139

(Significance codes: *** 0.001 , **0.01 , *0.05 . 0.01)

Table 5.2 represents both the stepwise regression model and the augmented model. The null hypothesis that interactions in the augmented model are jointly equal to zero was convincingly rejected. Thus, there was no support for the standards are uniform across households.

The household size has significant negative (though nonlinear) effects on welfare. This inverse relation between household size and the log per capita consumption is a common finding in the literature (Lanjouw and Ravallion, 1995; Lipton, 2001b).

Education variable emerge as a strong determinant of welfare. In both models the average years of schooling specified on its own have significant positive effects on per capita consumption. However, once the models have been augmented with interactions several interaction terms in schooling was found to be significant. For example, the marginal return to school was found to be increasing with household size, as well as decreasing with the number of the years in school.

The study found a strong positive significance effects on the average number of years in education for the family. The model of this study indicates strong positive effects on household if the family is educated.

This study found that family that owned land (for production) had a significant positive effect on per capita consumption of the household. Furthermore, the age of the household head shows the expected life cycle in the stepwise model increases with the age of the head, also the quadratic term of the age which is nonlinear shows a decline in the life cycle phenomenon of high earning capacity with greater experience and smoothing of consumption over life cycle.

Table 5.3: Augmented model adequacy

Models	Adjusted R^2
Stepwise	0.9642
Augmented	0.9701

The high value of adjusted R^2 shows that the variation in the dependent variables are explained by the explanatory variables and thus the model is quite adequate for the analysis, since 97% of the augmented model is explained by the regression (See Table 5.3).

5.3 Empirical studies of the Logistic Model

The logistic model predicts poverty directly because of the nature of the dependent variable. There are two things that need to be reiterated. First, the dependent variable takes values of 1 when the respondent is poor and 0 otherwise. This means in interpreting the estimation result it was important to remember that a positive coefficient means that the variables were correlated positively with the poor. Secondly, predicted value of the dependent variable is the probability of the observation to be poor.

A logit model was estimated to elicit the factors influencing welfare status of households. The model used welfare status of household as the dichotomous dependent variable. Poverty variable was defined on the basis of shortfall of food availability and/or access to a household during a year. This period was between harvesting season to the next harvest season which was approximately 163 days of the year (Mwita *et al.*, 2007).

In this study the following logistic regression model was employed to determine current poverty status :

$$P(Y_i = 1|X = x_i) = \pi = E\left[Y_i|x_1, x_2, \cdots, x_{p_i}\right], i = 1, 2, \cdots, n \tag{5.1}$$

Where $0 \leq \pi_i \leq 1$ and $\Pr(Y = 0) = 1 - \pi$

The logistic model defined as;

$$\pi_i = \frac{\exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ji}\right)}{1 + \exp\left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ji}\right)} \tag{5.2}$$

Where π_i was the probability that the household is poor, β_0 was the intercept term, β_j are coefficients and X_i for $i = 1, 2, \cdots, 13$ were the independent variables, the subscript i denoted the ith observation.

Let $Z_i = \beta_0 + \sum \beta_{ji}$, then $\pi_i = \dfrac{1}{1 + e^{-z}}$

As Z range from $-\infty$ to ∞, π_i ranges from 0 to 1 and π_i was non-linearly related to Z.

In estimation form, the model was;

$$L_i = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = Z_I = \beta_0 + \beta_j x_{ji} \tag{5.3}$$

Where L was the logit link, which shows the log odds in favour of the poverty status changes as the respective independent variable changes by unit.

The variables in this case are:

Y_i	Food security of household i ($1 =$ Poor, and $0 =$ Non-Poor)
X_1	Household size
X_2	Square of household size
X_3	Gender of household head ($1 =$ male, and $0 =$ female)
X_4	land size(acres)
X_5	Education of HH head ($1 =$Primary level and above,$0 =$No Education)
X_6	Age of HH head
X_7	Square of Age of HH head
X_8	per capita aggregate production (No. of Kgs)

In this study a thirteen-predictor logistic model was fitted to the data to test the relationship between the likelihood of a household being poor or non-poor. The logistic regression analysis was carried out by stepwise method, and the result show the predicted logit of the optimal model to be

$$\begin{aligned} Z_R &= -1.4721X_1 + 0.1398X_2 + 1.6905X_3 + 0.0358X_4 + 0.0781X_5 + 0.3796X_6 \\ &\quad -0.0059X_7 - 0.3659X_8 \end{aligned} \tag{5.4}$$

According to the model, the log of the odds of a household being poor was negatively related to size of the household ($p = 0.01$). This was in conformity with former studies. Paddy (2003), noted that household size was negatively correlated to food security and Deaton and Paxson (1995) found that food requirement increased in relation to the number of persons in household. The non-linear component of the household size was positively correlated to poverty, this is common finding in the literature (Lanjouw and Ravallion, 1995; Lipton, 2001a).

In this study, it was establish that the log of odds of the gender of the head of the household was positively related to the poverty ($p = 0.05$). This according to literature Anyanwu (1997) and a survey of food deprivation by gender household had 52% prevalence of undernourished compound to 48% for female leased households. In particular, it has been shown in many countries that poverty was higher in female headed households than in households headed by men. Recent examples of these findings are Gang *et al.* (2002) for the case of India; Anyanwu (2010) for Nigeria and Serumaga and Naude (2002) for South Africa. All of these authors found that poverty was higher in female headed households.

The age of the household head shows the expected life expectancy. In model ,household living standard increases with the age of the household head upto the optimal age of around 60 years,but decreases with the quadratic term which was significant $p = 0.05$.This consistent with higher earning with greater experience.

There was a strong inter-generational effect on education. Parental education had a strong positive correlation on household welfare.The results compare favorably with the GOK (2007). The poverty estimates are not directly comparable, given that different poverty lines, equivalence scale, time and data set are employed in estimating the headcount poverty index.

Food production was expected to be increased extensively through expansion of areas under utilization. The model indicates that land size increased food security with 42% even though ($p > 0.05$).

Based on the model of the study, the log odds of land size was positively related to poverty ($p = 0.05$). In other words, the larger the size of land the higher the production. As for the production (kg) of the household, the log of the odds indicates that a unit increase of food production resulted in an increase of the household by 1.7 of food poverty status of the household, with ($p > 0.05$).

Table 5.4: Predictors

Predictors	β	$SE(\beta)$	z	p-value	e^{β}(Odds ratios)
Size of HH (numbers)	-1.4721	0.0907	-1.603	0.1090·	0.2294
Square of household size	0.1398	0.080982	1.726	0.0844·	1.1500
Gender of HH head (1-male, 0-female)	1.6905	0.8790	1.923	0.0545·	1.0560
Land size(acres)	0.0358	0.2449	0.146	0.8836	2.4196
Education of HH head (1 =Primary level and above,0 =No Education)	0.0781	0.1054	0.741	0.4587	1.5820
Age of HH head	0.3796	0.1768	0.1768	0.0318 *	1.0000
Square of Age of HH head	-0.0059	0.0026	-2.231	0.0257 *	1.0260
per capita aggregate production (No. of Kgs)	-0.3659	0.1671	-2.4891	0.0139*	1.0140

5.3.1 Evaluation of the logistic regression model

Logistic regression model was used to correctly predict the category of an outcome for individual households. In the Stepwise regression framework, all independent variables the model were included in the analysis either by adding onto the model step by step and at each step the model fit was tested or alternatively by including all variables and eliminating one by one as appropriate. Coefficients of the variables were tested for significance and in the case of forward linear regression, they were added if significant while in the case of backward linear regression they were eliminated if found insignificant. Wald's test and likelihood-ratio test are used to test for significance while the Hosmer-Lemshow test was used to determine goodness-of-fit of the model.

The evaluations illustrated below are based on model Equation 3.29. The overall model evaluation was said to provide a better fit to the data if it demonstrated an improvement over the intercept only model (also called the null model) (Hosmer and Lemeshow, 2000). An intercept only model served as a good baseline because it contained no predictors. According to the model of this, all observations would be predicted to belong in the largest outcome category. An improvement over this baseline was examined by using three inferential statistical tests.

Table 5.5: Statistical inference table Statistical test

Statistics Test	χ^2	df	p
Likelihood ratio test	9.0353	5	0.0486
Wald test	4.0456	5	0.0544
Goodness of fit test			
Hosmer–Lemeshow	9.6702	5	0.7418

The statistical significance of individual regression coefficient that is, (β's) were tested using the Wald chi-square statistic. According to Table 5.5, the variables are significant predictors of poverty ($p < 0.05$).

Goodness-of-fit statistics assess the fit of a model against actual values. The inferential goodness-of-fit test in the Hosmer- Lemeshow (H-L) test that yield a $\chi^2_{(5)}$ of 9.6702 and was insignificant ($p > 0.05$). Suggesting that the model fitted well to the data.

Table 5.6: 95% confidence interval for one unit change in X_t

Size of HH (Number)	$-3.7940, 0.08354$
Square of household size	$0.0002, 0.3473$
Gender of HH head (1-Male, 0-Female)	$0.1099, 3.8349$
Land size	$-0.4561, 0.3260$
Education of HH head (1 =Primary level and above,0 =No Education	$-1.0770, 2.5661$
Age of HH head	$0.0930, 0.8789$
Square of Age of HH head	$-0.0120, 0.0014$
Per capita aggregate production (kg)	$-0.7457, -0.1033$

The full model of the study was:

$$\begin{aligned} Z_F = &-2.5237X_1 + 0.2237X_2 + 2.4270X_3 + 0.0358X_4 \\ &+ 0.7810X_5 + 0.6391X_6 - 0.00932X_7 - 0.3834X_8 - 2.076X_{11} - 0.0024X_{13} \end{aligned} \tag{5.5}$$

Hypothesis tested was as follows:

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \cdots \beta_{10}$$
$$H_A : \beta_j \neq 0$$

The reduced model was:

$$\begin{aligned} Z_R &= -1.4721X_1 + 0.1398X_2 + 1.6905X_3 + 0.0358X_4 + 0.0781X_5 + 0.3796X_6 \\ &\quad -0.0059X_7 - 0.3659X_8 \end{aligned} \quad (5.6)$$

From the Table 5.5 the likelihood ratio test indicates that the reduced model was better than the full model with a test statistic of 9.0353 and p-value of 0.0486 ($p<0.05$) and thus the study rejected the null hypothesis and concluded that reduced model was better than the full model.

5.4 Assets as a Measure of Poverty

5.4.1 Housing conditions

5.4.1.1 Roofing Material as measure of poverty

Table 5.7: Roofing Materials Variables

Roofing material	Frequency	Percentage
Corrugated Iron Sheets	3029	68.6
Tiles	49	1.1
Concrete	33	.7
Asbestos Sheets	94	2.1
Grass	1192	27.0
Makuti	3	.07
Tin	1	0
Other	1	0

Finding from household data collected during the National population census in 2009 KNBS (2013), established that majority of respondent represented by 68.6 % stay in corrugated iron sheet houses, followed by glass thatched houses at 27.0 %. There were 1.1 % houses roofed with tiles, another 2.1 % with asbestos and other with 0.07% roofed by other material. This factor may not give a good indicator of poverty but if looked from the perspective of the whole house building material it was found this indicator can be able to give some indication of poverty.

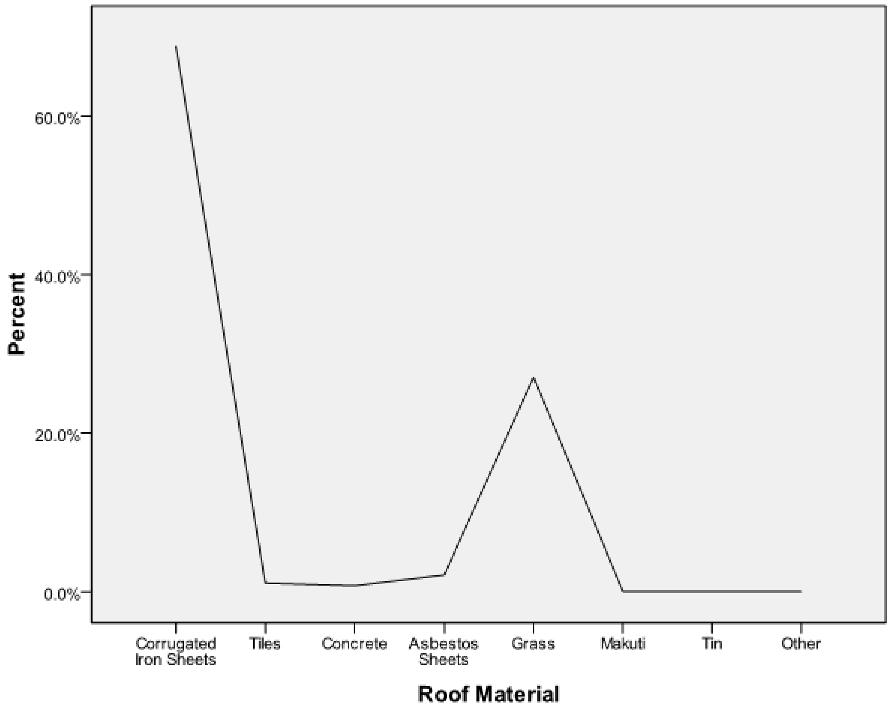

Figure 5.3: Roofing Material

5.4.1.2 Wall material as a measure of poverty

Table 5.8: Materials used for building walls

Wall Materials	Frequency	Percentage
Stone	81	1.8
Brick/Block	803	18.2
Mud/Wood	2746	62.2
Mud/Cement	679	15.4
Wood Only	5	0.1
Corrugated Iron Sheets	40	0.9
Grass/Reeds	4	0.1
Tin	1	0
Other	43	1.0

The study also found that majority of houses had wall made of mud and wood which represents 62.2%, 18.2% are made of bricks, 15.4% had walls made of mud and cement and others 3% have walls made with other materials like timber and stone which indicates that combined with roofing materials, the household in the region were not good dwelling areas for it was expect the inhabitants to be inclined to jigger manifestation.

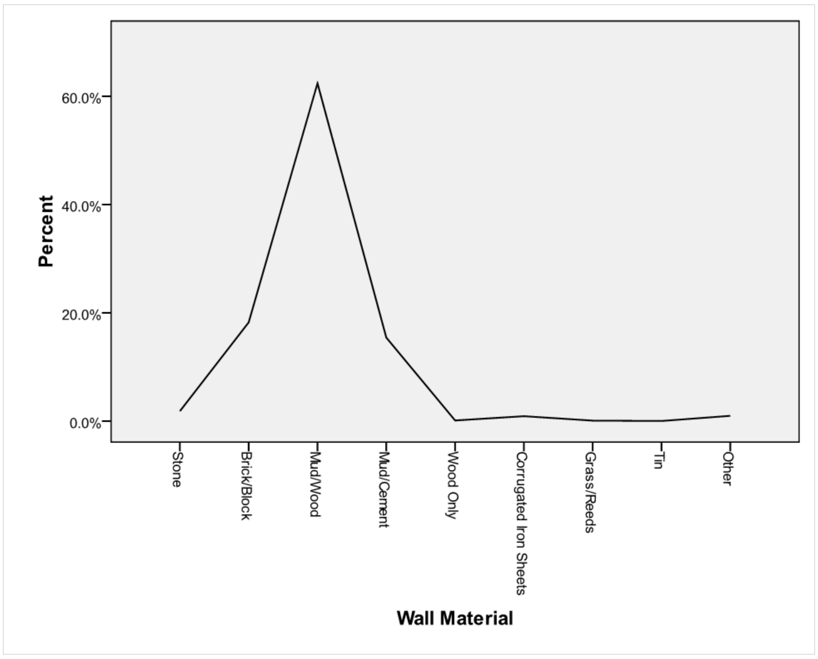

Figure 5.4: Wall Material

5.4.1.3 Main water sources as a measure of poverty

Table 5.9: Water sources

Water sources	Frequency	Percentage
Pond	93	2.1
Dam	93	2.1
Lake	755	16.8
Stream	2017	44.8
Borehole	744	16.5
Piped into dwelling	73	1.6
Piped	408	9.1
Jabia	3	0.1
Rain/Harvested	31	0.7
Water Vendor	195	4.3

Access to clean water was an important indicator of a health family. From the study many of the household used water from the stream, 44.8% which many time was not treated and as such water borne diseases were rampant in these areas. This is followed by the uptake of lake water which was contaminated with affluent and chemicals as in the case lake Victoria basin.

5.4.2 Poverty against livestock

5.4.2.1 Poverty against Indigenous cattle

The findings on correlation between poverty and the raring of indigenous cattle in the region, $\chi^2 = 155.835$ with 13 degrees of freedom at $p < 0.05$. Since the p-value was less than 0.05, the null hypothesis was rejected and concluded that there was statistical significance association between poverty and the rearing of the indigenous cattle in the region. The sample size requirement for chi-squared test of independence was satisfied.

5.4.2.2 Poverty against Goat

The findings on correlation between poverty and the raring of goat in the region, $\chi^2 = 85.213$ with 11 degrees of freedom at $p < 0.05$. The relationship between poverty and goat rearing was also statistically significant.

5.4.2.3 Poverty against Sheep

The findings on correlation between poverty and the raring of sheep in the region, $\chi^2 = 30.444$ with 8 degrees of freedom at $p < 0.05$. In the region there existed a relationship between poverty and sheep rearing which was statistically significant.

5.5 Forecasting Food Production and Consumption

In the context of food security,which was given high priority by the governments either at national and regional (county) levels, the issues and prospects concerning the three components of food security access, availability and stability, as well as the question of food quality needs to be considered in the proper perspective (FAO, 2006).

The projections in this section are not statements of what will happen, but of what might happen given the assumption and methods used. The reference case projection are business as usual and forecasts, given known technological and demographical trends and current laws and regulations. Thus these projections provide a policy neutral starting point that can be used for analysis of National and regional food requirements and policy initiatives.

5.5.1 Assumptions and methodology

From the Equation 3.13 and Equation 3.17, assuming that $\varepsilon \sim N_n\left(0, \sigma^2 I_n\right)$, leads to the familiar ordinary least squares (ols) estimator of β.

$$\beta_{ols} = \left(X'X\right)^{-1} X'Y \tag{5.7}$$

with covariance matrix,

$$V\left(\beta_{ols}\right) = \sigma^2 \left(X'X\right)^{-1} \tag{5.8}$$

It was assumed more generally that $\varepsilon \sim N_n(0,\Sigma)$ where the error covariance matrix Σ was symmetric and positive definite.

Different diagonal entries in Σ correspond to non-constant error variances, while non-zero off-diagonal entries correspond to correlated errors.

Suppose, that Σ was known, then the log likelihood for the model was;

$$log_eL(\beta) = \frac{-n}{2}log_e 2\pi - \frac{1}{2}log_e(det(\Sigma)) - \frac{1}{2}(y - x\beta)'\Sigma^{-1}(y - x\beta) \quad (5.9)$$

which was maximized by the generalized least squares (GLS) estimator of β,

$$\beta_{GLS} = \left(X'\Sigma^{-1}X\right)^{-1}X\Sigma^{-1}Y \quad (5.10)$$

with covariance matrix

$$V(\beta_{GLS}) = \left(X'\Sigma^{-1}X\right)^{-1} \quad (5.11)$$

For example,when Σ was a diagonal matrix of (generally) unequal errors variances, then β_{GLS} was just the weighted least squares estimator. Since the error-covariance matrix Σ was not known and must be estimated from the data along with the regression coefficients β with a suitably restrictive parametrization of Σ, the model was estimated by maximum likelihood (Fahrmeir and Kaufmann, 1986).

5.5.2 Empirical Results

This study was based on data collected in 2007 by researchers at Vicres (Mwita *et al.*, 2007). The data on food balance sheet was collected from the household under study ,the agriculture offices and the district statistical office for the period 31st March 2003 to 1st April 2007.

Food Balance Sheet for each of the sampled households was compiled. Although main source of food was through own production, the following variables were used in the Balance Sheet as additions to or subtractions from own production of three main district food crops at household level: food purchases (+), food received as aid (+), post harvest food losses (-), cereals used for seed (-), food marketed (-). The transformation was made from available food in kilograms to total available calories for each household by using the standard conversion factors, as they are in kinds (FAO, 2002). Secondly, the food available at household level calculated in step one was used to calculate calories available per person per day for each household. That was the Household calorific acquisition (per capita calories) was obtained by controlling time and household size. In that case, total calories available in the household to be consumed was divided by one hundred and fifty, with the latter being the average duration from planting up to harvest of all food available, and got the average Kilo calories per day. This was then divided by the number of members of household to get the average kilo calories to be consumed per day per person in the household. Thirdly, 1683 kilo calories per person per day was used as a measure of calories required (demanded) to maintain body-weight and perform a sedentary light physical activity taking account of age and sex structure of the Kenya population (KNBS, 2007c).

In this section, projections are made for the four years of between 31st March 2003 to 1st April 2007. The analysis was to estimate future consumption and production of the selected food crops. The result of analysis are shown in the Figures 5.5-5.13. The self-sufficiency ratio for the selected food crops in the region Kisumu, Kuria and Siaya are also shown.

Maize remained the dominant cereal in the region over the projection period. The decline in the self-sufficiency of maize in the region imply the region imports a lot of maize from outside, either in the neighbouring counties. The region having very good soils needs some education or agricultural incentive in order to improve on maize production.

The production and consumption of sorghum in the region with a self-sufficiency averaging 57.80% in Kisumu, 51.67% in Siaya and 57.80% in Kuria in 2007 shows a projected increase in the next four years.

The production and consumption of beans in the region which was the main protein contributor in the region as observed from the FBS tables, in Kisumu the production of beans seems to be doing well and the projection of the same for a period of four years shows an increase in production and consumption of beans in comparison to the other regions.

Year	*Production* (1000 *tons*)	*consumption* (1,000 *tons*)	Self sufficiency %
2003-2004	37.80	49.30	76.67
2004-2005	40.99	50.42	81.30
2005-2006	43.20	45.47	95.01
2006-2007	45.26	52.44	86.31
2007-2008	43.33	50.47	85.85
2008-2009	42.69	50.72	84.16
2009-2010	41.83	50.94	82.11
2010-2011	46.76	51.13	79.71

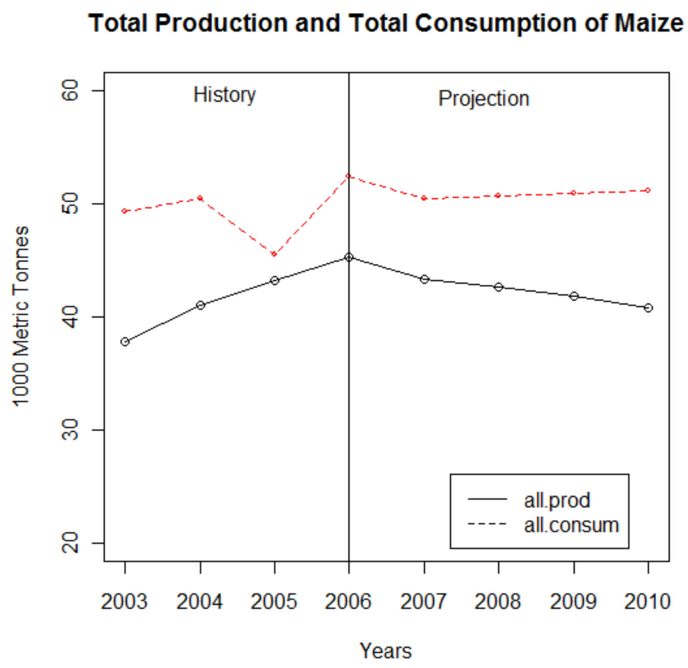

Figure 5.5: Total Production and Consumption of Maize (Kisumu Region)

Year	*Production* (1000 *tons*)	*consumption* (1,000 *tons*)	Self sufficiency %
2003-2004	18.60	21.28	87.41
2004-2005	21.15	24.06	87.90
2005-2006	25.00	28.65	87,26
2006-2007	18.03	22.56	79.92
2007-2008	23.94	27.57	86.83
2008-2009	23.58	27.71	85.09
2009-2010	23.12	27.83	83.07
2010-2011	22.56	27.94	80.74

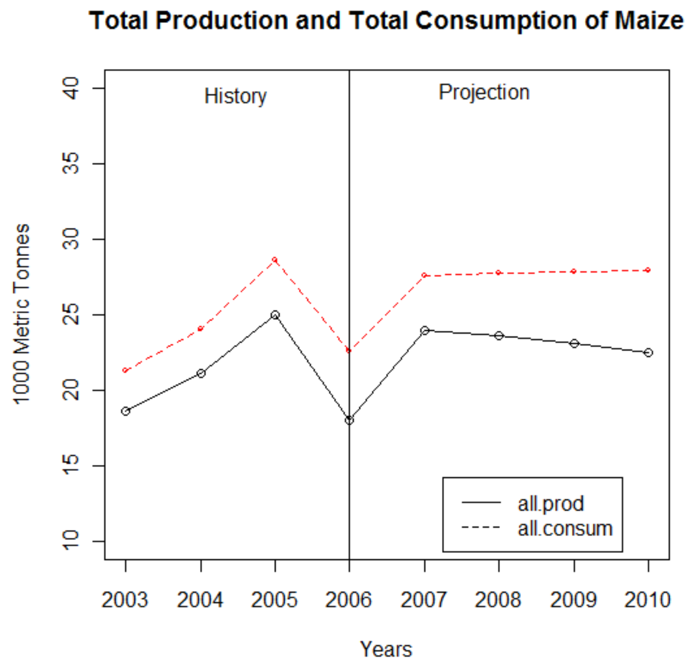

Figure 5.6: Total Production and Consumption of Maize (Kuria Region)

Year	*Production* (1000 *tons*)	*consumption* (1,000 *tons*)	Self sufficiency %
2003-2004	15.60	34.32	45.45
2004-2005	13.83	26.20	52.78
2005-2006	14.29	30.40	47.01
2006-2007	15.74	38.66	40.71
2007-2008	15.07	37.20	40.51
2008-2009	14.85	37.39	39.71
2009-2010	14.55	37.56	38.74
2010-2011	14.18	37.69	37.62

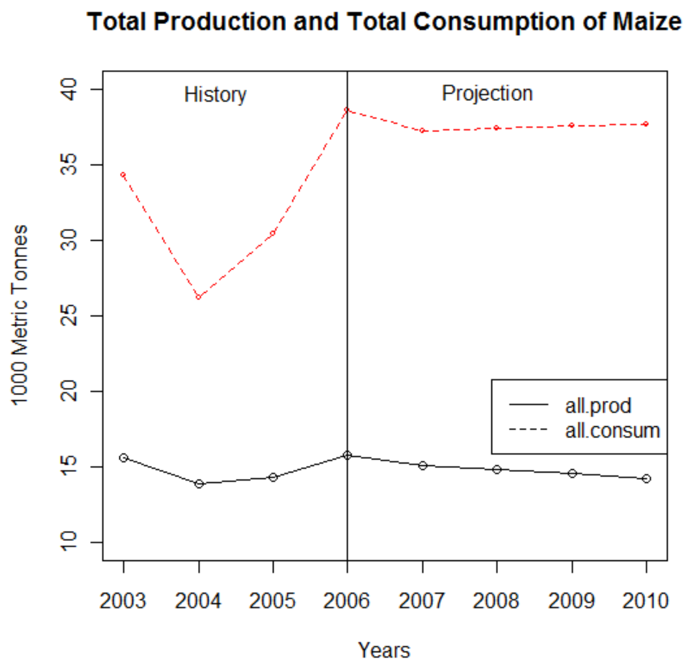

Figure 5.7: Total Production and Consumption of Maize (Siaya Region)

Year	*Production* (1000 *tons*)	*consumption* (1,000 *tons*)	Self sufficiency %
2003-2004	7.27	11.00	66.09
2004-2005	8,26	14.65	56.34
2005-2006	9.17	11.46	80.02
2006-2007	6.46	11.84	54.56
2007-2008	8.78	14.10	62.27
2008-2009	8.68	14.17	61.26
2009-2010	8.48	14.23	59.59
2010-2011	8.26	14.29	57.80

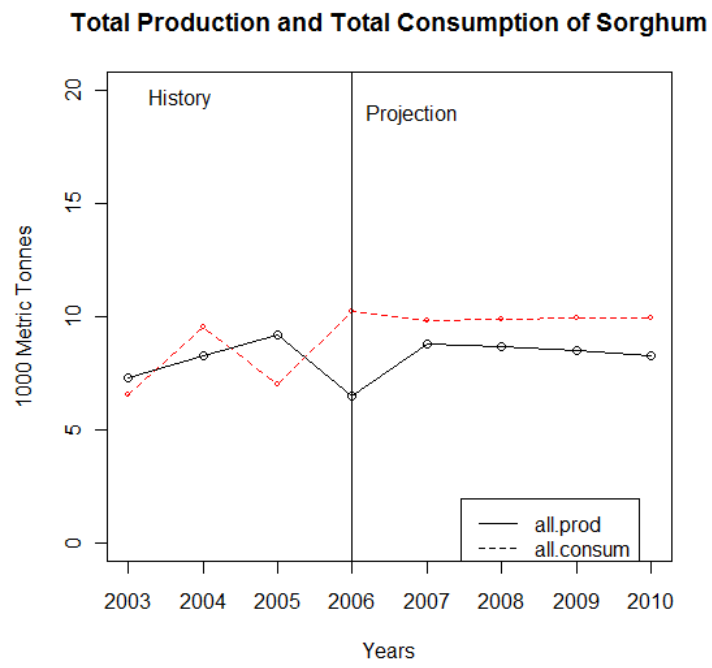

Figure 5.8: Total Production and Consumption of Sorghum (Kisumu Region)

Year	Production (1000 tons)	consumption (1,000 tons)	Self sufficiency %
2003-2004	7.27	11.00	66.09
2004-2005	8,26	14.65	56.34
2005-2006	9.17	11.46	80.02
2006-2007	6.46	11.84	54.56
2007-2008	8.78	14.10	62.27
2008-2009	8.68	14.17	61.26
2009-2010	8.48	14.23	59.59
2010-2011	8.26	14.29	57.80

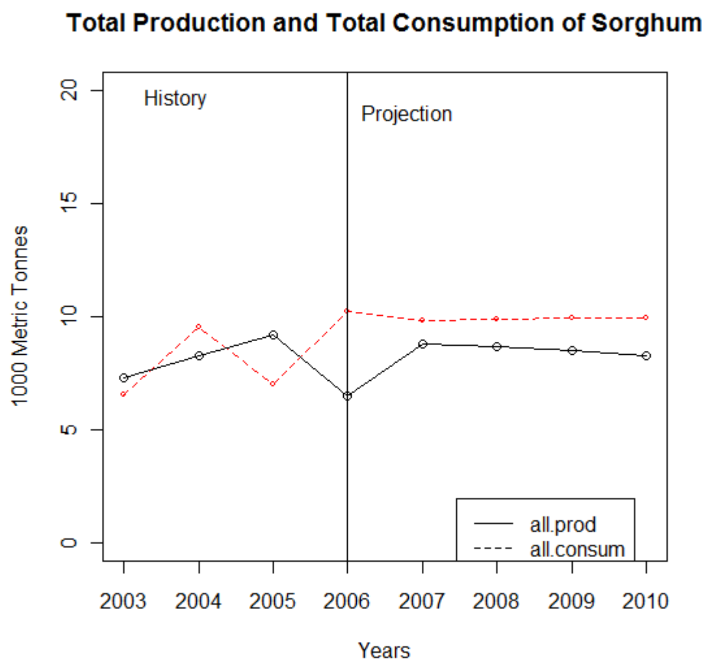

Figure 5.9: Total Production and Consumption of Sorghum (Kuria Region)

Year	Production (1000 tons)	consumption (1,000 tons)	Self sufficiency %
2003-2004	7.41	13.24	55.96
2004-2005	6.61	9.56	69.14
2005-2006	4.46	7.01	63.62
2006-2007	5.88	10.20	57.65
2007-2008	7.09	12.74	55.65
2008-2009	6.99	12.81	54.57
2009-2010	6.85	12.86	53.26
2010-2011	6.67	12.91	51.67

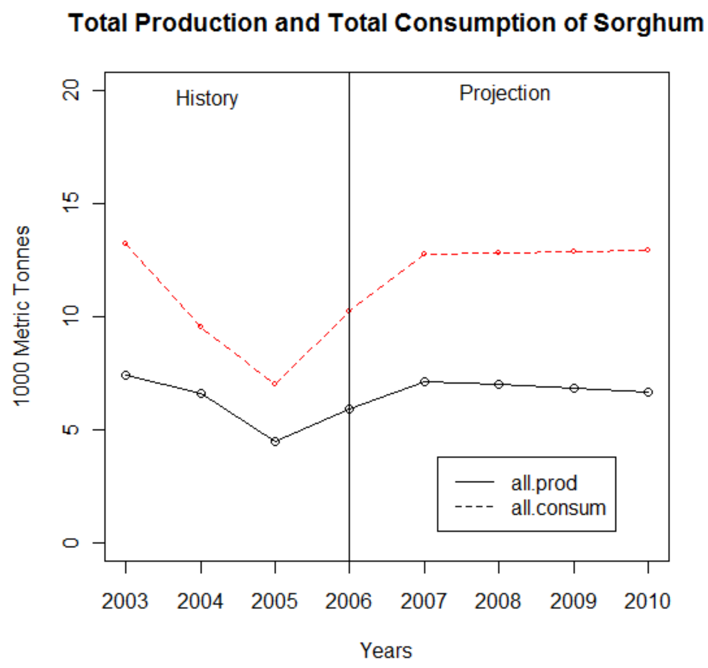

Figure 5.10: Total Production and Consumption of Sorghum (Siaya Region)

Year	Production (1000 tons)	consumption (1,000 tons)	Self sufficiency %
2003-2004	11.50	13.19	87.18
2004-2005	12.90	13.77	93.68
2005-2006	10.13	14.45	70.10
2006-2007	9.15	11.90	76.89
2007-2008	12.35	13.86	89.10
2008-2009	12.16	13.93	87.29
2009-2010	11.93	13.99	85.27
2010-2011	11.61	14.04	79.30

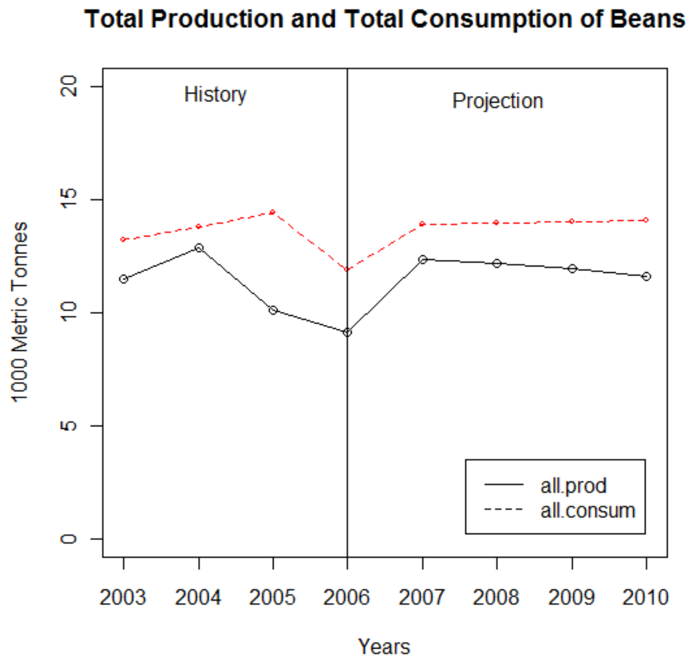

Figure 5.11: Total Production and Consumption of Beans (Kisumu Region)

Year	Production (1000 *tons*)	*consumption* (1,000 *tons*)	Self sufficiency %
2003-2004	2.23	5.09	43.81
2004-2005	1.46	3.75	38.93
2005-2006	1.1	2.33	43.78
2006-2007	2.55	4.53	56.04
2007-2008	2.44	4.90	49.79
2008-2009	2.41	4.92	48.98
2009-2010	2.36	4.94	47.77
2010-2011	2.30	4.96	46.37

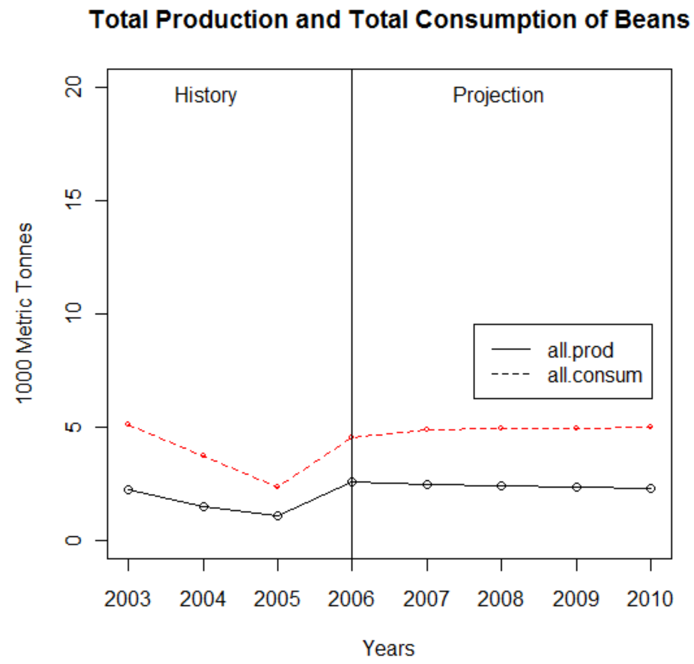

Figure 5.12: Total Production and Consumption of Beans (Kuria Region)

Year	*Production* (1000 *tons*)	*consumption* (1,000 *tons*)	Self sufficiency %
2003-2004	6.55	14.80	44.26
2004-2005	6.93	18.60	37.26
2005-2006	5.75	12.55	45.82
2006-2007	8.83	13.56	65.12
2007-2008	8.45	17.09	49.44
2008-2009	8.34	17.99	46.36
2009-2010	8.16	18.09	45.10
2010-2011	7.95	18.14	43.82

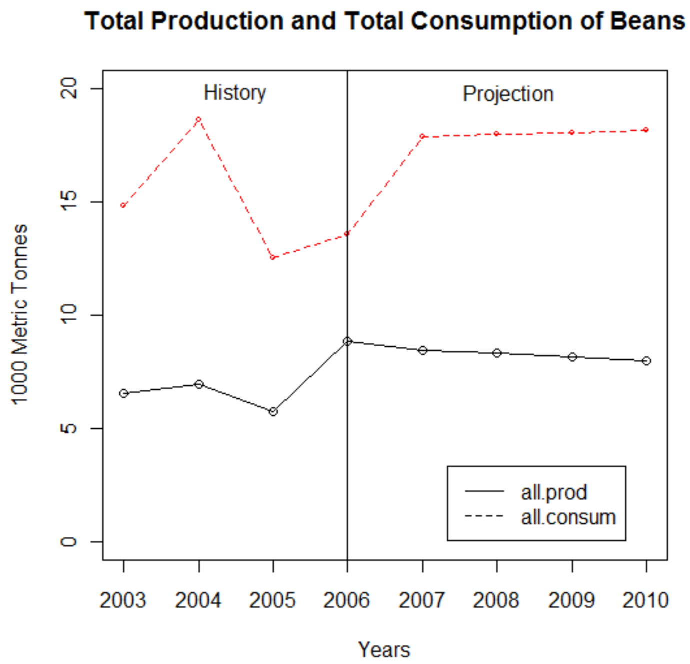

Figure 5.13: Total Production and Consumption of Beans (Siaya Region)

Domestic grain production in a region is a major source of food supply and fluctuations in production are a major cause of instability in food availability. The before and after forecasting on food production show that food grain production increased by 15.3% in Kisumu, by 4% in Kuria and by 6.9% in Siaya from 2003-2007 as depicted in the tables 5.10, 5.11 and 5.12.

Table 5.10: Food Balance sheet for the Kisumu region in 2003-2007

Variable	2003-2004	2004-2005	2005-2006	2006-2007	Growth Rate (%)
Production of cereals (x1000 metric tonnes)	41.82	43.68	53.86	62.65	15.3
Less 15% post-harvest loss	6.273	6.552	8.079	9.395	
Less 6% seed	2.5092	2.6208	3.2316	3.7591	
Net production	33.078	34.5072	42.5494	49.4975	15.3
Import and stock changes	41.64	46.16	71.71	75.48	
Total grain available	74.6778	80.6672	114.2594	124.9775	20.8
Total supply	74.6778	80.6672	114.2594	124.9775	
Total population	557.980	547.384	552.446	562.149	
Food requirement @ 225kgs/HHs	125.545	123.164	124.300	126.4831	3.2
Food balance	-51.4677	-42.4942	-10.0406	-1.5056	
Self sufficiency ratio	0.2632	0.2802	0.3422	0.3913	14.9
Food availability ratio	0.5900	0.6550	0.9192	0.9889	20.9

Source Mwita *et al.* (2007)

This was because production of early maturing cereals such as millet and sorghum was encouraged in the region, this could also be coupled by the growth of institutional infrastructure and positive shift of policy during this period. The highest growth was noted in 2006-2007 which could be due to the better economic times the country was experiencing that period (KNBS, 2007a).

Based on the available data, a FBS was computed for the period 2003-2007 using 2100 calories of food per day as the minimum nutritional requirements. This was equivalent to 225 kilograms of cereals per person per annum. FBS are the principle tools used for calculating national food security which was used to determine the expected food deficit or surplus, the necessary food requirements (Frankenberger, 1992).

Table 5.10 reveals that food requirement in Kisumu grows at 3.2%, while food availability increased by 20.8%. As a result, the food deficit declined by 14.9%. The Self Sufficiency Ratio (SSR) which was measured as a ratio of the sum of the net production in relation to domestic utilization more so, the food availability ratio which is the ratio of the food supply to the requirement has increased by 20.9%. As the figures from the Food Balance Sheet reveals, the food availability and the district food self-sufficiency was increasing over the period 2003-2006. Food security at the regional level that is, self-reliance in food at the regional level does not necessarily mean food security at the district level.

Table 5.11 shows that food requirement in Kuria increased by 2.3% while food availability increased by 2.0%. As a result food deficit declined by 2.2%. The SSR decreased by 1.9% and the food availability increased by 0.1%. The food sufficiency magnitude expressed the magnitude of production in relation to domestic utilization.

As the figure from the food balance sheet reveals, the food availability and the district food sufficiency growth was minimal in the period 2003-2007.

Table 5.11: Food Balance sheet for the Kuria region in 2003-2007

Variable	2003-2004	2004-2005	2005-2006	2006-2007	Growth Rate (%)
Production of cereals (x1000 metric tonnes)	25.71	25.428	30.195	24.683	4
Less 15% post-harvest loss	3.8565	3.8142	4.52925	3.70245	
Less 6% seed	1.5426	1.52568	1.8117	1.48098	
Net production	20.3109	20.08812	23.85405	19.49962	4.9
Import and stock changes	5.28	5.50000	5.980	6.550	
Total grain available	25.5909	25.588	29.83405	26.04962	2
Total supply	25.5909	25.588	29.83405	26.04962	
Total population	184.721	189.123	193.380	197.882	
Food requirement @ 225kgs/HHs	41.5622	42.5526	43.5105	44.52345	2.3
Food balance	-15.9713	-16.96455	-13.6764	-18.47383	-2.203
Self sufficiency ratio	0.4887	0.47720	0.5482	0.43796	-1.8807
Food availability ratio	0.6157	0.6013	0.68567	0.585076	0.1

source:Mwita *et al.* (2007)

Table 5.12 reveals that food requirement in Siaya decreased by 0.40% while food availability decreased by 0.5%. As a result the food deficit declined by 9.6%. The SSR increased by 6.8%. The self-sufficiency magnitude ratio expressed as magnitude in production which grows by 6.9% in relation to the domestic availability. The food availability ratio which was the ratio of the food supply to the requirement had also decreased by 0.1%.

Table 5.12: Food Balance sheet for the Siaya region in 2003-2007

Variable	2003-2004	2004-2005	2005-2006	2006-2007	Growth Rate (%)
Production of cereals (x1000 metric tonnes)	40.79	48.97	44.49	52.49	6.86%
Less 15% post-harvest loss	6.1185	7.3455	6.693	3.1494	
Less 6% seed	2.4474	2.9382	2.6772	3.1494	
Net production	32.2241	38.6863	35.2498	41.4671	6.86%
Import and stock changes	39.11	66.89	54.80	41.37	
Total grain available	79.40	105.5763	90.0498	82.8371	
Total supply	79.90	105.5763	90.0498	82.8371	-0.506%
Total population	494,728	492,826	493,326	488,034	
Food requirement @ 225kgs/HHs	111.3138	110.88585	110.99835	109.80765	-0.3978%
Food balance	-31.4138	-5.30955	-20.94855	-26.97085	-9.58%
Self sufficiency ratio	0.2895	0.3664	0.3157	0.3776	6.75%
Food availability ratio	0.7177	0.9548	0.81127	0.75438	-0.134%

source:Mwita *et al.* (2007)

As the figures for the FBS reveals, the availability and the district food self sufficiency decreased in the period 2003-2007. Food security at the district level needs to be improved through policy mitigation on decreasing in population. Introduction of fast growing crop such as millet and sorghum should be emphasized.

In conclusion, even though the FBS provides useful information regarding trends in food availability they are often too aggregated to detect patterns of food deficit or vulnerability in a given region or district. FBS is not usually drawn upon a disaggregated basis to detect differences across district or region Frankenberger (1992).

CHAPTER SIX

SUMMARY, CONCLUSION AND RECOMMENDATIONS

This chapter draws the summary, conclusions and gives recommendations on the way forward on the methodology and policy.

6.1 Summary

The first objective of the study sought to come up with a threshold for assessing the poverty level of a household. The study finding shows that many different statistical indices of poverty can be used as a threshold of poverty. As noted, the FGT indices are the commonly used (Mwabu *et al.*, 2000; Mukui, 1994; Geda *et al.*, 2001), but the finding using the FEI gave better results with minimal limitations.

In line with the results obtained in the poverty profile, the negative sign of the logistic regression parameter for household size indicates the existence of a direct relationship between poverty and household size. Also, it was observed that an increase of one member to the size of the household, increases the log of odds of being poor by 22.6 percent. This negative effect of household size upon poverty coincides with the findings obtained for the case of India (Gang *et al.*, 2002), Nigeria (Anyanwu, 2010) and South Africa (Serumaga and Naude, 2002)

The second objective of the study sought to investigate two statistical models to determine the best model for poverty studies. The study established that the logistic model has higher and better results for determining the characteristics of poverty in an area for it had higher prediction rate than ordinary least squares regression model. Findings by Pohlman and Leitner (2003) also concluded that both models can be used to test relationships with a binary criterion. However, logistic regression is superior to Ordinary Least Squares at predicting the probability of an attribute and should be the model of choice for that application.

Most African countries, government regard the provision of formal housing, water and sanitation services as naturally urban services, but as the countries develop it would not be amiss for the rural population to strive towards having piped water, flush toilets and good housing characteristics. It was possible that important changes may take place in the economic situation of many households, but the asset indices may remain unchanged. That being the case then we cannot asset can not be a measure for short or medium term social welfare of a household.

The third objective of the study sought to implement a mathematical model for predicting the FBS data for mitigation against food security. The study modelled the FBS data using the logistic model and obtained projections which could be the best starting point for the region in checking food crops status of the region. The results collaborated very much with (Dawoud, 2005).

6.2 Conclusion

These findings indicated a correlation between better farming conditions and lower poverty, which was something that can be useful to emphasize as an initial step in poverty reduction for the rural areas. This in turn can reduce the inefficiency in the rural agricultural sector through, for example, infrastructural improvements, such as higher accessibility to water, better schools and other social amenities . However, the correlation between farming conditions and poverty does not necessarily give much information about the causal effect because a farmer that has more wealth can also have better farming conditions and therefore lower poverty.

There was lack of consensus on how to measure poverty in general, even though poverty indices and poverty profiles are increasingly being used as guides in targeting resources to reduce poverty. An allocation that is efficient according to one methodology may yield unacceptable results when a different methodology is applied.

Results from poverty studies were also sensitive to the choice of poverty line (the means of identifying the poor) and poverty measure (the measure obtained when aggregating incomes or expenditures of households below poverty line). Since the choices are typically at the discretion of the analyst, this has given rise to the suggestion that the results obtained are not robust. Potentially different results could be obtained by the choice of a different poverty line or measure. Moreover, few conclusions can be drawn if poverty trends differ substantially when different poverty measures are applied or the position of the poverty line is changed.

Analysts have tried to overcome the problem by employing a number of poverty lines/ measures but this only partially overcomes the problem since it may still be possible to obtain different results by the choice of another poverty line. Thus, what is ideally needed is an approach that is robust to the choice of poverty line.

Most of the studies in poverty in Kenya have used the summary measures to determine the extent and level. The robustness of poverty measures using summary measures such as means and variances can be compromised by errors in living standards data, unknown differences between households at similar consumption levels, uncertainties and arbitrariness in both the poverty lines and the precise poverty measure.

Further the poverty profile by different socioeconomic characteristics shows that larger households are not necessarily poorer in the rural areas in fact the reverse effect is apparent. However, this does not imply that a larger household size reduces poverty, as there is a cost related to supporting more household members than less.

The results presented in this study should enable policy makers to explore the challenges involved in achieving local and international development goals especially the goal of reducing poverty by half by the year 2015 as indicated in the Millennium Development Goals (MDGs). The achievement of this goal would allow Kenya to make irreversible progress towards a better life. In the formulation of these goals, justifying the targeting of any poverty alleviation policy and programs to a category of people or region, ascertaining and monitoring the impact of such programs and projects form time to time, the need of robust evidence is inevitable.

As mentioned in this study, poverty is more prevalent in the rural areas and although not mentioned in this study, poverty is also prevalent among specific categories of people such as female headed households and pastoralists. Eradication should, as a matter of priority be focused in these areas.

The analysis found that there is a steady increase in future consumption of most selected food items on the one hand. Production for most food items, on the other hand, appears to be leveling off. If this continues, the region will become more and more dependent on other regions to meet its food requirement. To reverse the situation will be very difficult. There has to be a drastic change in the positioning of the agricultural sector. What is needed is not merely a stabilisation programme, but structural changes to the entire sector. Taking both consumption and production together, the region is expected to be getting further and further away from being self-sufficient in its food production,

6.3 Recommendation

To increase the number of food secure households in the region, it is recommended that; fertilizer and pesticides/herbicide should be made accessible to household farmers: household should be encouraged to exercise family planning to reduce the household size. Extensive agriculture services should be intensified to enable household farmer adopt new agriculture methods and technology for efficient land use.

6.3.1 Future Research Areas

The issue of food balance sheet within the local communication should be researched further to give a precise warning times in case of poverty within the areas.

The nutritional effects to the households should be well researched further and the soil texture in these areas and the rainfall patterns in the three districts.

Future studies should look at the relationship between the respondents economic activities and poverty levels in the region.

REFERENCES

Adrich, J. (1997). Statistical science project. Technical report. Retrieved from: euclid.org.

Agresti, A. (2002). *Introduction to Categorical Data Analysis*. John Wiley & Sons Inc, NewYork.

Amemiya, T. (1985). *Advanced Econometric*. Havard University Press, Cambridge.

Anyanwu, D. J. (2010). Poverty in Nigeria: A gendered analysis. *The African Statistical Journal*, 11:1–15.

Anyanwu, J. D. (1997). Poverty in Nigeria: Concepts, measurements and determinants. Technical report, Proceedings of the Nigerian Economic Society (NES).

Beer, M. (2001). *Asymptotic Properies of the Maximum Likelihood Estimator in Dichotomous Logistic Regression Models*. PhD thesis, University of Fribourg, Switzerland.

Bell, C. S. (1995). What is poverty? *American Journal of Economics and Sociology*, 54(2):161–162.

Booysen, F., Van der Berg, S., Burger, R., Von Maltitz, M., and du Rand, G. (2008). Using an asset index to assess trends in poverty in seven sub-saharan african countries. *World Development*, 36(6):1113 –1130.

Bouis, H. and Haddad, L. (1990). Agricultural commercialization, nutrition and the rural poor. Technical report, International Food Policy Research Institute (IFPRI), Boulder, Colorado.

Cameron, C. A. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge University Press.

Chen, M. R. . S. (2007). Absolute poverty measures for the developing world. *Journal of Developement Economics*, 82:1–42.

Chen, S. and Ravallion, M. (2002). How have the world's poorest fared since the early 1980's. Technical report, World Bank, Washington, DC.

Dalaker, J. and Naifeh, M. (2005). Alternative poverty estimates in the united states. Technical report, U.S. Government Printing Office, Washington, DC.

Dalton, H. (1920). The measurement of the inequality of incomes. *The Economic Journal*, 30(119):348–361.

Dawoud, S. (2005). *An Analysis of Food Consumption Patterns in Egypt*. PhD thesis, University of Kiev, Ukraine.

Deaton, A. and Paxson, C. (1995). Measuring poverty among the elderly. Technical report, National Bureau of Economic Research.

Deutsch, R. C. (2007). *Benchmark Analysis for Two Predictor Variables*. PhD thesis, University of South Carolina, South Carolina.

Diaz-Bonilla, E., Thomas, M., and Robinson, S. (2003). Trade, food security and wto negotiations: Some reflections on boxes and their content. *Agricultural Trade and Poverty: Making Policy Analysis Count*, 30(2):59–104.

El-Obeid, A. E., Johnson, S. R., Smith, L. C., and Jensen, H. H. (1999). Commodity supply and use balances and food availability. In *Food security: New solutions for the 21st century. Proceedings from the Symposium honoring the Tenth Anniversary of the World Food Prize*, pages 35–72, Iowa State. Iowa State University Press.

Elbers, C., Lanjouw, J., and Lanjouw, P. (2001). Welfare in villages and towns: micro-level estimation of poverty and inequality. Technical report, World Bank, Washington, DC.

Ercelawn, A. (1991). Absolute poverty as risk of hunger: Norms, incidence and intensity for rural and urban pakistan. Technical report, Applied Economics Research Centre, University of Karachi.

Fahrmeir, I. and Kaufmann, H. (1985). Consistency and asymptotic normality of maximum likelihood estimator in generalized linear models. *Annals of Statistics*, 13:342–368.

Fahrmeir, I. and Kaufmann, H. (1986). Asymptotic inference in discrete response models. *Statistische Hefte*, 27:179–205.

Fan, J., Farmen, M., and Gijbels, J. (1998). Local maximum likelihood estimator and inference. *Stat. Soc. Ser. B Stat. Methodol.*, 60(3):591–608.

FAO (1985). Energy and protein requirements,report of a joint fao/who ad hoc expert consultation. Technical Report 724, Food and Agriculture Organization of the United Nations, Geneva,WHO.

FAO (1996). The sixth world food survey. Technical report, Food and Agriculture Organization of the United Nations, Rome.

FAO (2002). The state of food insecurity in the world. Technical report, Food and Agriculture Organization of the United Nations, Rome, Italy.

FAO (2003). Food insecurity and vulnerability information and mapping systems. Technical report, Food and Agriculture Organization of the United Nations, Rome, FAO.

FAO (2006). The state of food insecurity in the world, 2006. Technical report, Food and Agriculture Organization of the United Nations, Rome, Italy.

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52(3):761–766.

Frankenberger, T. R. (1992). Indicators and data collection methods for assessing household food security. Technical report, IFAD, Rome, Italy, IFAD, 107 Via del Serafico.

Gang, N. I., Sen, K., and Yun, M. S. (2002). Caste, ethnicity and poverty in rural india. Departmental Working Paper 200634, New Brunswick Rutgers University, Department of Economics, New Jersey.

Geda, A., de Jong, N., Mwabu, G., and Kimenyi, M. S. (2001). Determinants of poverty inkenya: Household-level analysis. Discussion Paper Series 9, The Kenya Institute for Public Policy Research and Analysis (KIPPRA), Nairobi.

Geda, A., de Jong, N., Mwabu, G., and Kimenyi, M. S. (2005). Determinants of poverty in kenya household-level analysis. Economics Working Papers 200544., The Kenya Institute for Public Policy Research and Analysis, Nairobi.

GTZ (2006). Food security policy advice. Technical report, German Agency for Technical Cooperation, Nairobi.

Gillin, E. (2000). Food balance sheets. Technical report, FAO, Rome. Italy. FAOSTAT Database.

Gourienx, C. and Monfort, A. (1981). Asymptotic properties of the maximum likelihood estimation in dichotomous logit models. *Econometric*, 7:83–97.

GOK (2007). Kenya integrated household budget survey 2005/06. Technical report, Kenya National Bureau of Statistics, Ministry of Finance and Planning. Government printer, Nairobi.

Greer, J. and Thorbecke, E. (1986a). Food poverty profile applied in kenya smallholders. *Economic Development and Cultural Change*, 35(1):115–141.

Greer, J. and Thorbecke, E. (1986b). A methodology for measuring food poverty applied in kenya. *Journal of Developement Economics*, Vol. 24:pp. 59–74.

Grigg, D. (1993). International variation in food consumption in the 1980. *Annals of American Geography*, 78(340):251–266.

Grigg, D. (1996). The starchy staples in world food consumption. *Annals of American Geography*, 86(3):412–432.

Hagenaars, A. and Vos, K. (1988). The definition and measurement of poverty. *The Journal of Human Resources*, 23(2):211–221.

Hentschel, J., Lanjouw, J., Lanjouw, P., and Poggi, J. (2000). Combining census and survey data to trace the spatial dimensions of poverty,a case study of ecuador. *World Bank Economic Review*, 14(1):147–165.

Hoddinott, T. (1996). Operationalizing household security and development strategies. Technical report, International Food Policy Research Institute, Technical Cavidelizen,No.1 Washington, DC.

Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression 2*, volume 2. John Wiley & Sons, New York.

Hurst, C. (2004). *Social inequality: Forms, causes and consequences.* Person Education Inc, New York, NY.

Joanes, D. N. (1994). Reject inference applied to logistic regression for credit scoring. *J. Math. Appl. Bus. Ind*, 5:35–43.

Jordan, B. (1996). *A theory of poverty and social exclusion*. Polity Press, Cambridge, MA.

Kabubo-Mariara, J. (2007). Poverty and policy in Kenya. Proceedings of a national workshop, University of Nairobi Press, Nairobi,Kenya.

KNBS (2007a). Basic report on well-being in Kenya. Technical report, Kenya National Bureau of Statistics, Regal Press Kenya, Nairobi, Kenya.

KNBS (2007b). Basic report on well being in Kenya based on kenya integrated household budgets survey, 2005/2006. Technical report, Kenya National Bureau of Statistics, Nairobi.

KNBS (2007c). Geographic dimension of well being in kenya. Technical report, Kenya National Bureau of Statistics, Nairobi.

KNBS (2013). Basic report on well-being in kenya. Technical report, Kenya National Bureau of Statistics.

Kotler, P., Roberto, N., and Leisner, T. (2006). Alleviating poverty: A macro/micro marketing perspective. *Journal of Macromarketing*, 26(3):233–238.

Laderchi, C., Saith, R., and Stewart, F. (2003). Does it matter that we do not agree on the definition of poverty: A comparison of four approaches. *Oxford Development Studies*, 31(3):233–274.

Lanjouw, P. and Ravallion, M. (1995). Poverty and household size. *The Economic Journal*, 105(433):1415–1434.

Lehmann, E. L. and G.Casella (1998). *Theory of point estimation*. Springer ,New York, New York.

Lipton, M. (2001a). Challenges to meet: food and nutrition security in the new millennium. *Nutrition Society*, 60(2):203–214.

Lipton, M. (2001b). Reviving global poverty reduction: what role for genetically modified plants? *Journal of International Development*, 13(7):823–846.

Maddala, G. (1983). *Limited Dependent Variables and Quantitative Variables in Econometrics*. Cambridge University Press., Cambridge.

Madden, D. (2000). Relative or absolute poverty lines: A new approach. *Review of Income and Wealth*, 46(2):18–99.

Malaba, J. (2006). Poverty measurement and gender: Zimbabwe experience. Technical report, United Nations Secretariat, Department of Economic and Social Affairs, United Nations Secretariat, Department of Economic and Social Affairs, Statistics Division. ESA /STAT /AC. 122/ 14.2.

Maxwell, S. and Smith, M. (1992). Household food security a conceptual review. in household food security: Concepts, indicators, and measurements: A technical review. Technical report, UNICEF and IFAD, New York and Rome.

McFadden, L. (1974). *Conditional logit analysis of qualitative choice behaviour*. Frontiers in Econometrics. Academic, New York.

Mukui, J. T. (1994). Kenya: Poverty profiles, 1982-92. Technical report, Ministry of Finance and Planning, Nairobi, mimeo.

Mwabu, G., Masai, W., Gesami, R., Kirimi, J., Munene, F., Chemengich, M., and Mariara, J. (2000). Poverty in kenya; profiles and determinants. Technical report, University of Nairobi and Ministry of Finance & Planning, Mimeo, Nairobi.

Mwabu, G., Mwangi, S., Kimalu, P., Nafula, N., and Manda, D. (2002). Predicting household poverty: A methodology note with a kenyan example. Discussion Paper 12, Kenya Institute for Policy Research and Analysis, Nairobi.

Mwita, P., Masanja, V., Muyanja, C., and Odhiambo, R. (2007). Mathematical techniques for food crops balance sheet and food security indicators in lake victoria watershed. Technical report, Inter-University Council of East Africa, Arusha.

OECD (2000). Waging the global war on poverty strate giesand case studies. Retrieved from: http://www.oecd.org/document.

Oiro, M. W., Mwabu, G., and Manda, D. (2004). Poverty and unemployment in kenya. Discussion Paper 33, KIPPRA, Nairobi.

Oyugi, L. N. (2000). The determinants of poverty in kenya. Master's thesis, Department of Economics, University of Nairobi. MA Thesis, Department of Economics ,University of Nairobi.

Paddy, F. (2003). Gender differentials in land ownership and their impact on household food security: A case study of masaka. Master's thesis, Uni–Hohenheim, Uganda. Retrieved from: http:www.troz.uni–hohenheim.de/research/thesis/MSCAES/paddy.pdf.

Pastor-Barriuso, R., Guallar, E., and Coresh, J. (1998). Use of two-segmented logistic regression to estimate change of point in epidemological studies. *American Journal of Epidemiology*, 148(7):631–642.

Pastor-Barriuso, R., Guallar, E., and Coresh, J. (2003). Transition model for change point estimation on logistic regression. *Statist. Med*, 22:1141–1162.

Pohlman, J. T. and Leitner, D. W. (2003). A comparison of ordinary least squares and logistic regression. *The Ohio Journal of Science*, 103(5):118–125.

Prescott, N. and Pradhan, M. (1997). A poverty profile of cambodia. Discussion Paper 373, World Bank, Washington, DC.

Rank, M. (2004). *One Nation Underprivileged: Why American Poverty Affects Us All*. Oxford Press., New York, NY.

Ravallion, M. (1996a). How well can method substitute for data? five experiments in poverty analysis. Technical Report 11(2), The World Bank Research Observer, Washington, DC.

Ravallion, M. (1996b). *Issues in measuring and modeling poverty*. Number 1615. World Bank-free PDF, Washington, DC.

Ravallion, M. (1998). Poverty lines in theory and practice. LSMS Working Paper 133, The World Bank, Washington DC, The World Bank.

Ravallion, M. (2008). Global poverty reassessed: A reply to reddy. Technical report, International Policy Centre for Inclusive Growth.

Ravallion, M. and Bidani, B. (1994). How robust is a poverty profile? *World Bank Economic Review*, 8(1):75–102.

GoK (1998). First report on poverty in kenya: Incidence and depth of poverty. Technical report, Ministry of Finance and Planning, Nairobi.

Schiller, R. B. (2008). *The Economics of poverty and discrimination*. Peason Prentice Hall, Upper Saddle River, NJ.

Sen, A. (1979). Personal utilities and public judgements, or whats wrong with welfare economics? *The Economic Journal*, 89:537–580.

Sen, A. (1981). *Poverty and Famines*. Clarendon press, Oxford,UK.

Sen, A. (1985). *Commodity capabilities*. Amsterdam, North-Holland.

Sen, A. (1986). Food,economics, and entitlements. *Lloyd's Bank Review*, 160:1–20.

Serumaga, P. Z. and Naude, W. (2002). The determinants of rural and urban household poverty in the north west province of south africa. *Development Southern Africa*, 19(4):561–572.

Shifa, M. R. . N. (2009). Consistency of the maximum likelihood estimator in logistic regression model:a different approach. *Journal of Statistics*, 16:1–11.

Smith, L. C. (2009). Can fao's measure of chronic undernourishment be strengthened? *Food Security*, 1(1):5–7.

Svedberg, P. (2011). How many people are malnourished? *Annual review of nutrition*, 31:263–283.

Svedberg, P. (2012). Reforming or replacing the public distribution system with cash transfers? *Economic & Political Weekly*, 47(7):53–62.

WB (1986). Poverty and hunger: Issues and orphans for food security in developing countries. Technical report, World Bank, Washington, DC.

WB (1997). Voices of the poor; poverty and social capital in tanzania. Technical report, World Bank, Washington, D.C.

WB (2000). World development report. Technical report, Oxford University Press, Oxford.

WB (2001). Overview understanding, measuring and overcoming poverty. Technical report, World Bank. Retrieved from http://web.worldbank.org/.

WB (2006). World development indicators. Technical report, World Bank, Washington DC.

WB (2007). Accelerating equitable growth; poverty reduction and economic management unit africa region. Technical Report 38662, World Bank, Washington, DC. Ethiopia Country Economic - Momerandum.

WB (2010). World development indicators 2010. Technical Report License: CC BY 3.0 IGO, World Bank, Washington, DC. Retrieved from: https://openknowledge.worldbank.org/handle/10986/4373.

APPENDICES

Appendix 1: Kenya selected indicators, 2008

Population, total (millions)	38.77
Population growth (annual %)	2.6
Surface area (sq. km) (thousands)	580.4
GNI, Atlas method (current US$) (billions)	28.42
GNI per capita, Atlas method (current US$)	730
GNI, PPP (current international $) (billions)	60.32
GNI per capita, PPP (current international $)	1.560
Life expectancy at birth, total (years)	54
Fertility rate, total (births per woman)	4.9
Primary completion rate, total (% of relevant age group)	80
GDP (current US$) (billions)	30.35
GDP growth (annual %)	1.7
Inflation, GDP deflator (annual %)	13.1
Agriculture, value added (% of GDP)	27

Source: World Development Indicators, 2010

Appendix 2: An example of FBS

Domestic supply(1,000 tons)

Product	Production	Imports	Stock changes	Exports	Total	···
Cereals	388,032	18,305	6,982	1,828	415,147	···
Wheat	110,569	9,352	1,030	972	121,923	···
Maize	···	···	···	···	···	···

Domestic Utilization(1,000 tons)

Feed	Seed	Processing	Waste	Other uses	···
129,557	9,385	11,057	23,089	972	···
4,005	4,800	1,770	5,105	469	···
···	···	···	···	···	···

Per Capita Availability(1,000 tons)

Total Food	Kg per year	kcal	Protein	Fat
242059	192.7	1,671	36.8	5.1
106243	84.3	616	17.5	2.5
···	···	···	···	···

Appendix 3: The Meaning of Odds

Odds represent the relative frequency with which different outcomes occur. Odds are sometimes expressed as a ratio of the form a:b. For example, offs of 3:1 in favor of the first outcome means that the first outcome occurs 3 times for each single occurrence of the second outcome. Similarly, offs of 5:2 means that first outcome occurs 5 times for each 2 occurrences of the second outcome. Odds are directly related to probabilities and can be translated back and forth using these relations: $Probability = \frac{a}{(a+b)}$ when odds are expressed as a:b, or $Probability = Odds(1+Odds)$ when odds are expressed in decimal form (e.g. 3:1 becomes 3:0 and 5:2 becomes 2:5). $Odds = \frac{Probability}{(1-Probability)}$ Some examples : The probability of rolling a 1 with a true die is $\frac{1}{6} = .1667$. The odds in favor of rolling a one are . or 1:5. Viewed the other way, the odds against rolling a 1 with a true die are 5:1, or 5.0. The odds in favor of not drawing a face card from a standard deck of playing cards (i.e. drawing a card other than a Jack, Queen, King or Ace) are 36:16 or 9:4. The corresponding probability is $\frac{9}{(9+4)} = \frac{9}{13} = .6923$.

Some examples : The probability of rolling a 1 with a true die is $\frac{1}{6} = .1667$. The odds in favor of rolling a one are . or 1:5. Viewed the other way, the odds against rolling a 1 with a true die are 5:1, or 5.0. The odds in favor of not drawing a face card from a standard deck of playing cards (i.e. drawing a card other than a Jack, Queen, King or Ace) are 36:16 or 9:4. The corresponding probability is $\frac{9}{(9+4)} = \frac{9}{13} = .6923$

Appendix 4: Food Crops Balance Sheets (2003-2007)

The FBS measures total food supply for domestic use and is a measure at the most aggregate level of the food distribution system. It includes food available for both thehousehold and non-household sectors. The NHS measures total food consumed or acquired at the household level. The FBSs compile information on all food available for human consumption as a residual of supply minus non-food use during a period of one year. The data quantities used are food produced and imports, from which are subtracted food exports, feed to animals, seed and other non-food uses, such as biofuel. The net amount of food available for human consumption was usually expressed on a per person per day basis. This amount was obtained by dividing the net food quantity by the country's population size and the number of days of the reference year. This per person per day amount of food was also expressed in dietary energy and macronutrients values. The food quantity of each food item was converted into macronutrient values using the public health food composition table (PHFT). It was then aggregated to give the nutrient consumption of dietary energy, protein and fat at the food item level.

The main purpose of the FBS is to estimate the overall food supply meant for human consumption in the country, on a yearly basis. Annual FBSs over a period of years show trends in the overall national food supply and reveal changes that may have occurred in the type of food consumed. For example, such changes may include differences in the composition of the diet and the impact of agricultural and food policies in terms of food production, trade and use. Data sources come from the statistical system within the institutions of the Ministry of Agriculture and Trade and Statistics and Kenya National Bureau of Statistics (KNBS). Data analysis is constrained by the fact that basic data is collected from different sources, which are often inconsistent, incomplete and unreliable.

NHS food data is collected from families during a short-term period and relates to consumed food in contrast with the raw food crops of the FBS. NHS reveals the composition of the diet and, if available for regular periods, supplies the trend analysis of the food commodity items. These are useful indicators for analysing the food supply patterns from the FBS.

The following tables show the FBS for the three districts from 2003 to 2007.

Food Crops Balance Sheet of Kisumu District April 1st 2003 - 31st March 2004

KISUMU DISTRICT, POPULATION /CENSUS 557,980 1st April 2003 - 31th March 2004															
	DOMESTIC SUPPLY (1000 METRIC TONNES)					DOMESTIC UTILIZATION (1000 METRIC TONNES)						PER CAPITAL SUPPLY PER DAY			
	Production	Imports	Stock changes	Exports	Total district supply	Feed	Seed	Processed	Waste	Other utilization	Food	Food per person per year	Calorie	Protein	Fat
PRODUCTS	1000 METRIC TONNES											Kg	Units	Grams	Grams
Maize	21.80	28.64	1.00	0.00	51.44	0.00	0.00	0.00	2.00	0.00	49.44	98.02	956.00	25.51	11.55
Sorghum	7.27	2.73	2.00	0.00	12.00	0.00	0.00	0.00	1.00	0.00	11.00	21.81	212.72	5.68	2.57
Millet	1.25	2.38	3.00	0.00	6.62	0.00	0.00	0.00	2.00	0.00	4.62	9.16	89.39	2.39	1.08
Beans	11.50	0.89	1.00	0.00	13.39	0.00	0.00	0.00	0.20	0.00	13.19	26.14	254.97	6.80	3.08
TOTAL												155.13	1513.09	40.38	18.28

Food Crops Balance Sheet of Kisumu District April 1st 2004- 31st March 2005

KISUMU DISTRICT, POPULATION /CENSUS 547,384 1st April 2004 - 31th March 2005															
	DOMESTIC SUPPLY (1000 METRIC TONNES)					DOMESTIC UTILIZATION (1000 METRIC TONNES)						PER CAPITAL SUPPLY PER DAY			
	Production	Imports	Stock changes	Exports	Total district supply	Feed	Seed	Processed	Waste	Other utilization	Food	Food per person per year	Calorie	Protein	Fat
PRODUCTS	1000 METRIC TONNES											Kg	Units	Grams	Grams
Maize	20.99	29.45	2.00	0.00	52.44	0.00	0.00	0.00	2.00	0.00	50.44	100.00	975.34	26.03	11.78
Sorghum	8.26	17.39	1.00	0.00	26.65	0.00	0.00	0.00	2.00	0.00	24.65	48.87	476.69	12.72	5.76
millet	1.53	2.35	3.00	0.00	6.87	0.00	0.00	0.00	3.00	0.00	3.87	7.68	74.92	2.00	0.90
beans	12.90	0.87	0.10	0.00	13.87	0.00	0.00	0.00	0.10	100.00	13.77	27.30	266.31	7.11	3.22
TOTAL												183.86	1793.26	47.85	21.66

Food Crops Balance Sheet of Kisumu District April 1st 2005- 31st March 2006

KISUMU DISTRICT, POPULATION /CENSUS 552,446 1st April 2005 - 31th March 2006															
	DOMESTIC SUPPLY (1000 METRIC TONNES)					DOMESTIC UTILIZATION (1000 METRIC TONNES)						PER CAPITAL SUPPLY PER DAY			
	Production	Imports	Stock changes	Exports	Total district supply	Feed	Seed	Processed	Waste	Other utilization	Food	Food per person per year	Calorie	Protein	Fat
PRODUCTS	1000 METRIC TONNES											Kg	Units	Grams	Grams
Maize	21.20	29.24	3.00	0.00	53.44	0.00	0.00	0.00	8.00	0.00	45.44	90.09	878.65	23.45	10.61
Sorghum	9.17	8.29	2.00	0.00	19.46	0.00	0.00	0.00	2.00	0.00	17.46	34.62	337.67	9.01	4.08
millet	13.36	23.66	1.00	0.00	38.03	0.00	0.00	0.00	1.00	0.00	37.03	73.41	716.01	19.11	8.65
beans	0.10	0.90	0.20	0.00	1.20	0.00	0.00	0.00	0.10	0.10	1.10	2.18	21.27	0.57	0.26
TOTAL												200.30	1953.60	52.13	23.60

Food Crops Balance Sheet of Kisumu District April 1st 2006- 31st March 2007

KISUMU DISTRICT, POPULATION /CENSUS 562,149 1st April 2006 - 31th March 2007															
	DOMESTIC SUPPLY (1000 METRIC TONNES)					DOMESTIC UTILIZATION (1000 METRIC TONNES)						PER CAPITAL SUPPLY PER DAY			
	Production	Imports	Stock changes	Exports	Total district supply	Feed	Seed	Processed	Waste	Other utilization	Food	Food per person per year	Calorie	Protein	Fat
PRODUCTS	1000 METRIC TONNES											Kg	Units	Grams	Grams
Maize	21.26	29.17	8.00	0.00	58.44	0.00	0.00	0.00	6.00	0.00	52.44	103.97	1014.02	27.06	12.25
Sorghum	6.46	3.54	2.00	0.00	12.00	0.00	0.00	0.00	3.00	0.00	9.00	17.84	174.04	4.64	2.10
millet	22.33	22.77	8.00	0.00	53.10	0.00	0.00	0.00	8.00	0.00	45.10	89.41	872.10	23.27	10.53
beans	0.15	0.85	0.20	0.00	1.20	0.00	0.00	0.00	0.30	0.20	0.90	1.78	17.40	0.46	0.21
TOTAL												213.01	2077.57	55.44	25.09

Food Crops Balance Sheet of Kuria District April 1st 2003- 31st March 2004

KURIA DISTRICT, POPULATION /CENSUS 184,721 1st April 2003 - 31th March 2004															
	DOMESTIC SUPPLY (1000 METRIC TONNES)					DOMESTIC UTILIZATION (1000 METRIC TONNES)						PER CAPITAL SUPPLY PER DAY			
	Production	Imports	Stock changes	Exports	Total district supply	Feed	Seed	Processed	Waste	Other utilization	Food	Food per person per year	Calorie	Protein	Fat
PRODUCTS	1000 METRIC TONNES											Kg	Units	Grams	Grams
Maize	18.60	0.84	1.84	7.45	28.73	0.00	0.00	0.00	0.51	0.00	28.22	152.77	1490.04	39.76	18.00
Beans	2.23	0.92	0.67	1.40	5.22	0.00	0.00	0.00	0.13	0.00	5.09	27.58	269.02	7.18	3.25
Sorghum	4.88	0.34	0.84	0.67	6.70	0.00	0.00	0.00	0.16	0.00	6.54	35.42	345.48	9.22	3.20
TOTAL												215.77	2104.54	56.16	24.45

Food Crops Balance Sheet of Kuria District April 1st 2004- 31st March 2005

KURIA DISTRICT, POPULATION /CENSUS 189,123 1st April 2004 - 31th March 2005															
	DOMESTIC SUPPLY (1000 METRIC TONNES)					DOMESTIC UTILIZATION (1000 METRIC TONNES)						PER CAPITAL SUPPLY PER DAY			
	Production	Imports	Stock changes	Exports	Total district supply	Feed	Seed	Processed	Waste	Other utilization	Food	Food per person per year	Calorie	Protein	Fat
PRODUCTS	1000 METRIC TONNES											Kg	Units	Grams	Grams
Maize	21.15	0.67	3.04	7.04	31.9	0	0	0	0.475	0	31.425	166.162	1620.646	43.248	19.575
Beans	1.456	0.84	0.63	0.94	3.868	0	0	0	0.12	0	3.748	19.818	193.291	5.158	2.335
Sorghum	2.822	0.24	0.08	0.22	3.362	0	0	0	0.144	0	3.218	17.015	165.958	4.429	2.005
TOTAL												202.995	1979.895	52.834	23.914

Food Crops Balance Sheet of Kuria District April 1st 2005- 31st March 2006

KURIA DISTRICT, POPULATION /CENSUS 193,380 1st April 2005 - 31th March 2006															
	DOMESTIC SUPPLY (1000 METRIC TONNES)					DOMESTIC UTILIZATION (1000 METRIC TONNES)						PER CAPITAL SUPPLY PER DAY			
	Production	Imports	Stock changes	Exports	Total district supply	Feed	Seed	Processed	Waste	Other utilization	Food	Food per person per year	Calorie	Protein	Fat
PRODUCTS	1000 METRIC TONNES											Kg	Units	Grams	Grams
Maize	25	0.63	3.5	8	37.13	0	0	0	0.48	0	36.65	189.523	1848.500	49.328	22.327
Beans	11	0.87	0.65	0.98	13.5	0	0	0	0.18	0	13.32	68.880	671.815	41.894	3.963
Sorghum	4.095	0.22	0.11	0.265	4.69	0	0	0	0.19	0	4.5	23.270	226.965	6.439	2.104
TOTAL												281.673	2747.280	97.661	28.394

Food Crops Balance Sheet of Kuria District April 1st 2006- 31st March 2007

KURIA DISTRICT, POPULATION /CENSUS 197,882 1st April 2006 - 31th March 2007															
	DOMESTIC SUPPLY (1000 METRIC TONNES)					DOMESTIC UTILIZATION (1000 METRIC TONNES)						PER CAPITAL SUPPLY PER DAY			
	Production	Imports	Stock changes	Exports	Total district supply	Feed	Seed	Processed	Waste	Other utilization	Food	Food per person per year	Calorie	Protein	Fat
PRODUCTS	1000 METRIC TONNES											Kg	Units	Grams	Grams
Maize	18.025	0.87	4.15	6.5	29.545	0	0	0	0.48	0	29.065	146.880	1432.588	38.229	17.304
Beans	2.55	0.64	0.51	1.01	4.71	0	0	0	0.18	0	4.53	22.892	216.381	13.924	1.317
Sorghum	2.108	0.29	0.09	0.21	2.698	0	0	0	0.19	0	2.508	12.674	119.103	3.507	1.146
TOTAL												182.447	1768.071	55.660	19.767

Food Crops Balance Sheet of Siaya District April 1st 2003- 31st March 2004

SIAYA DISTRICT, POPULATION /CENSUS 494,728 1st April 2003- 31th March 2004															
	DOMESTIC SUPPLY (1000 METRIC TONNES)					DOMESTIC UTILIZATION (1000 METRIC TONNES)						PER CAPITAL SUPPLY PER DAY			
	Production	Imports	Stock changes	Exports	Total district supply	Feed	Seed	Processed	Waste	Other utilization	Food	Food per person per year	Calorie	Protein	Fat
PRODUCTS	1000 METRIC TONNES											Kg	Units	Grams	Grams
Maize	5.16	20.72	5.00	0.00	30.88	0.00	0.00	0.00	7.00	0.00	23.88	48.40	472.07	12.60	5.70
Sorghum	17.41	8.26	3.00	0.00	28.66	0.00	0.00	0.00	3.50	0.00	25.16	51.01	497.52	13.28	6.01
millet	1.67	1.68	0.30	0.00	3.65	0.00	0.00	0.00	0.25	0.40	3.40	6.90	67.32	1.80	0.81
beans	6.55	18.45	7.00	0.00	32.00	0.00	0.00	0.00	8.00	0.00	24.00	48.65	474.50	12.66	5.73
TOTAL												154.96	1511.40	40.33	18.26

Food Crops Balance Sheet of Siaya District April 1st 2004- 31st March 2005

SIAYA DISTRICT, POPULATION /CENSUS 492,836 1st April 2004- 31th March 2005															
	DOMESTIC SUPPLY (1000 METRIC TONNES)					DOMESTIC UTILIZATION (1000 METRIC TONNES)						PER CAPITAL SUPPLY PER DAY			
	Production	Imports	Stock changes	Exports	Total district supply	Feed	Seed	Processed	Waste	Other utilization	Food	Food per person per year	Calorie	Protein	Fat
PRODUCTS	1000 METRIC TONNES											Kg	Units	Grams	Grams
Maize	33.83	38.47	5.00	0.00	77.30	0.00	0.00	0.00	9.00	0.00	68.30	138.45	1350.34	36.03	16.31
Sorghum	6.61	3.40	6.00	0.00	16.00	0.00	0.00	0.00	7.00	0.00	9.00	18.24	177.94	4.75	2.15
millet	1.60	1.75	0.20	0.00	3.56	0.00	0.00	0.00	0.20	280.00	3.36	6.80	66.33	1.77	0.80
beans	6.93	18.07	4.00	0.00	29.00	0.00	0.00	0.00	0.40	0.00	28.60	57.97	565.44	15.09	6.83
TOTAL												221.47	2160.05	57.64	26.09

Food Crops Balance Sheet of Siaya District April 1st 2005- 31st March 2006

SIAYA DISTRICT, POPULATION /CENSUS 493,326 1st April 2005- 31th March 2006															
	DOMESTIC SUPPLY (1000 METRIC TONNES)					DOMESTIC UTILIZATION (1000 METRIC TONNES)						PER CAPITAL SUPPLY PER DAY			
	Production	Imports	Stock changes	Exports	Total district supply	Feed	Seed	Processed	Waste	Other utilization	Food	Food per person per year	Calorie	Protein	Fat
PRODUCTS	1000 METRIC TONNES											Kg	Units	Grams	Grams
Maize	34.29	38.11	5.00	0.00	77.40	0.00	0.00	0.00	7.00	0.00	70.40	142.70	1391.86	37.14	16.81
Sorghum	4.46	5.54	2.00	0.00	12.00	0.00	0.00	0.00	5.00	0.00	7.00	14.19	138.40	3.69	1.67
millet	0.12	2.00	0.40	0.00	2.52	0.00	0.00	0.00	0.10	0.13	2.42	4.90	47.81	1.28	0.58
beans	5.75	19.15	8.00	0.00	32.90	0.00	0.00	0.00	0.35	0.00	32.55	65.98	643.54	17.17	7.77
TOTAL												227.78	2221.60	59.28	26.83

Food Crops Balance Sheet of Siaya District April 1st 2006- 31st March 2007

SIAYA DISTRICT, POPULATION /CENSUS 488,034 1st April 2006 - 31th March 2007															
	DOMESTIC SUPPLY (1000 METRIC TONNES)					DOMESTIC UTILIZATION (1000 METRIC TONNES)						PER CAPITAL SUPPLY PER DAY			
	Production	Imports	Stock changes	Exports	Total district supply	Feed	Seed	Processed	Waste	Other utilization	Food	Food per person per year	Calorie	Protein	Fat
PRODUCTS	1000 METRIC TONNES											Kg	Units	Grams	Grams
Maize	15.74	24.92	4.00	0.00	44.66	0.00	0.00	0.00	6.00	0.00	38.66	78.36	764.24	20.39	9.23
Sorghum	5.88	4.12	1.00	0.00	11.00	0.00	0.00	0.00	8.00	0.00	3.00	6.08	59.31	1.58	0.72
millet	2.04	1.32	0.20	0.00	3.56	0.00	0.00	0.00	0.10	0.00	3.46	7.00	68.31	1.82	0.83
beans	18.83	15.31	0.50	0.00	34.64	0.00	0.00	0.00	0.58	0.00	34.06	69.04	673.39	17.97	8.13
TOTAL												160.48	1565.26	41.77	18.91

Appendix 5: Questionnaire

Deployed for Household Interviews

District: . . . . Sub-Location: . . . . Village: . . . .

Household Code: . . . . Name of Household head: . . . .

1. Type of Household

- Male Headed: ☐ Female headed: ☐

2. Demographic characteristic of Household

S.N	Name of Household Members	Sex 1. Male 2. Female	Level of Education 1. No formal Education 2. Formal	Main source of livelihood 1. Food 2. Production 3. Petty Trade 4. Daily labour 5. Livestock	Secondary source
1					
2					
3					
4					

3. Average land holding size (in hectare) ....

4. How food secured is your household:

- Number of month's that your household was food secured 5 years ago? . . . .
- Number of months that your household is food secured now? . . . .

5.What is the estimated average annual total household income?

- 5 years ago . . . .
- Now ? . . . .

6. Asset ownership of the household in number

- Oxen . . . .
- Sheep holding . . . .
- Goat holding . . . .

- Cows ....

7. Changes on crop production (yield)

7.1 What was your actual amount of production before 5 years?

- Millet ....
- Maize ....
- Sorghum ....
- Beans ....

7.2 What was your actual amount of production for the last harvest season?

- Millet ....
- Maize ....
- Sorghum ....
- Beans ....

Appendix 6: Sample FBS-TABLE

A sample Food balance sheet table as used in conducting the research

Products	DOMESTIC SUPPLY (1000 METRIC TONNES)					DOMESTIC UTILIZATION (1000 MT)						PER CAPITA SUPPLY			
	Production	Imports	Stock changes	Exports	Total supply	Feed	Seed	Processed	Waste	Other utilization	Food	Per year	Per day		
												Per food	Calorie	Protein	Fats
1000 METRIC TONNES												Kg.	Units	grams	grams

Notes:

1. Column 1: Products- here write the 3 top food crops in the region,
e.g., 1st top crop: Maize, 2nd top crop: Rice, 3rd top crop: Cassava
2. Data will be collected for 4 marketing years (1st April 2003 – 31st March 2004; 1st April 2004 – 31st March 2005; 1st April 2005 – 31st March 2006; 1st April 2006 – 31st March 2007)
3. For each district 4 forms will be filled, each form for a marketing year.

Appendix 7: Article I

Logistic regression is widely used as a popular model for the analysis of binary data with the areas of applications including physical, biomedical and behavioral sciences. In this study, the logistic regression model, as well as the maximum likelihood procedure for the estimation of its parameters, are introduced in detail.

IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 10, Issue 4 Ver. III (Jul-Aug. 2014), PP 57-68
www.iosrjournals.org

On the Estimation and Properties of Logistic Regression Parameters

[1]Anthony Ngunyi, [2]Peter Nyamuhanga Mwita, [2] Romanus O. Odhiambo

Abstract: *Logistic regression is widely used as a popular model for the analysis of binary data with the areas of applications including physical, biomedical and behavioral sciences. In this study, the logistic regression model, as well as the maximum likelihood procedure for the estimation of its parameters, are introduced in detail. The study has been necessited with the fact that authors looked at the simulation studies of the logistic models but did not test sensitivity of the normal plots. The fundamental assumption underlying classical results on the properties of MLE is that the stochastic law which determines the behaviour of the phenomenon investigated is known to lie within a specified parameter family of probability distribution (the model). This study focuses on investigating the asymptotic properties of maximum likelihood estimators for logistic regression models. More precisely, we show that the maximum likelihood estimators converge under conditions of fixed number of predictor variables to the real value of the parameters as the number of observations tends to infinity.We also show that the parameters estimates are normal in distribution by plotting the quantile plots and undertaking the Kolmogorov -Smirnov an the Shapiro-Wilks test for normality,where the result shows that the null hypothesis is to reject at 0.05% and conclude that parameters came from a normal distribution.*
Key Words: *Logistic, Asymptotic, Normality, MRA(Multiple Regression Analysis)*

I. Introduction

Regression analysis is one of the most useful and the most frequently used statistical methods [24, 3]. The aim of the regression methods is to describe the relationship between a response variable and one or more explanatory variables. Among the different regression models, logistic regression plays a particular role. The basic concept, however, is universal. The linear regression model is, under certain conditions, in many circumstances a valuable tool for quantifying the effects of several explanatory variables on one dependent continuous variable. For situations where the dependent variable is qualitative, however, other methods have been developed. One of these is the logistic regression model, which specifically covers the case of a binary (dichotomous) response. [6] discussed an overview of the development of the logistic regression model. He identifies three sources that had a profound impact on the model: applied mathematics, experimental statistics, and economic theory. [?] also provided details of the development on logistic regression in different areas. He states that, “Sir [5] introduced many statisticians to logistic regression through his 1958 article and 1970 book, “The Analysis of Binary Data”. However, logistic regression is widely used as a popular model for the analysis of binary data with the areas of applications including physical, biomedical, and behavioral sciences.

In this study, the logistic regression models, as well as the maximum likelihood procedure for the estimation of their parameters, are introduced in detail. Based on real data set, an attempt has been made to illustrate the application of the logistic regression model.

Simulation is used in the study since it involves construction of complicated integrals that do not exists in a closed form that can be evaluated. Simulation methods can be used to evaluate it to within acceptable degrees of approximation by estimating the expectation of the mean of a random sample.

II. Literature Review

The method of maximum likelihood is the estimation method used in the logistic regression models, however, two other methods have been and may still be used for estimating the coefficient . These methods are the least squares and the discriminant function analysis. The linear model approach of analysis of categorical data proposed by Grizzle et al.(1969) used estimaton based on NonLinear Weighted S(NLWS). They demostrated that logistic model can be handled by the method of maximum likelihood using an iterative reweighted least squares algorithm. The discriminant approach to estimation of the coefficients is of historical importance as popularized by [4]. [14] compared the two methods when the model is dichotomous and

[1] Dedan Kimathi University of Technology $(Kenya)$
P.O BOX 657-10100, Nyeri
[2] Jomo Kenyatta University of Agriculture & Technology $(Kenya)$
P.O BOX 62000-00200, Nairobi

concluded that the discriminat function was sensitive to the assumption of normality. In particular, the estimation of the coefficient for the nonnormal distributed variables are biased away from zero, when the coefficient is in fact different from zero. This implies that for the dichotomous independent variable the discrimnant function will overestimate the magnitude of the coefficient.

According to [13], the fact concerning the interpretability of the coefficients is the fundamental reason why logistic regression has proven such a powerful analytic tool for epidemiologic research. At least, this argumentation holds whenever the explanatory variables x are quantitative. [9] investigate the asymptotic properties of various discrete and qualitative response models (including logit model) and provided conditions under which the MLE has its usual asymptotic properties, that is, the p-vector β of coefficients of linear combinations (x, β) has to be estimated from a finite sample of n observations. The method of analysis of generalized linear models can be used since logistic models are sub-category [17].

[11] established that the maximum likelihood estimators are the best asympotically and strong consistent estimators of the logit model, other estimators have been suggested for logit model including the minimum ϕ divergent estimator which are generalization of maximum likelihood and are also consistent and asymptotically normal [20].

[25] discussed the inconsistency of the generalized method of moments estimator of qualitative models with random regressors and suggested a suitable modification in case of the probit and not the logit.

In the parameter estimation and inference in statistics, maximum likelihood has many optimal properties in estimation: sufficiency (complete information about the parameter of interest contained in its estimation); consistency (true parameter value that generated the data recovered asymptotically, i.e. data of sufficiently large samples); efficiency (lowest possible variance of parameter estimates achieved asymptotically) and parameterization invariance. The asymptotic normality of the maximum likelihood in logistic regression models are also found in [18] and [19]. [18] presents regularity conditions for a multinomial response model when the logit link is used. [19] presents regularity conditions that assure asymptotic normality for the logit link in binomial response models and further verifies that his conditions are equivalent to those of [18]. [7] discuss the asymptotic distribution of the MLE for constructing confidence intervals and conducting tests of hypotheses.[12] prove that the MLE is asymptotically normal in this setting as long as certain regularity conditions are satisfied

2.1 Logistic function

The function has been discussed by many reseachers like [10]. It is given by;

$$f(g) = \frac{exp(g)}{1 + exp(g)}$$
$$= \frac{1}{1 + exp(g)} \quad (1)$$

when modelling a bernoulli random variable with multivariables, one directly models the probabilities of group membership, as follows;

$$P(Y = 1 \mid X = x) = \frac{1}{1 + exp\left(-\left(\beta_0 + \sum_{j=1}^{d} x_{ij}\beta_j\right)\right)} \quad (2)$$

where g in 1 is given by

$$g(X;\beta) = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_d X_{1d} \quad (3)$$

To illustrate, the applicability of the logistic function, the bold curve in the figure 0 shows that the logistic function puts more weight on the tails than the normal distribution.

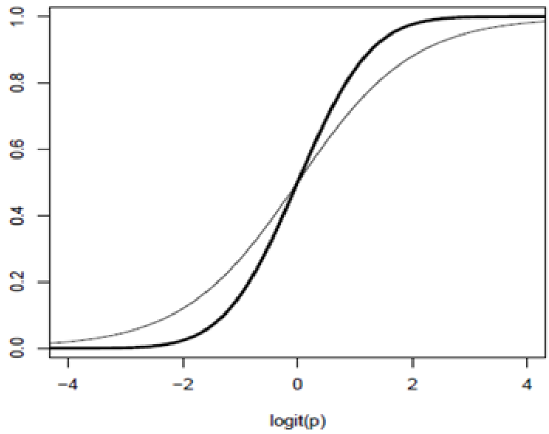

Figure 1: Standardized Normal and Logistic CDF's

The logistic model is bounded between zero and one, this property estimates the possibility of getting estimated or predicted probabilities outside this range which would not make sense. Also with a proper transformation, one can get a linear model from the logistic function. [10] uses the logit function of the Bernoulli distributed response variable. Transforming 2 as in [10] we have ;

$$Logit[P(Y=1 \mid X=x)] = log_e \frac{P(Y=1 \mid X=x)}{1-P(Y=1 \mid X=x)}$$

$$= log_e \left\{ \frac{1+exp\left(\beta_0 + \sum_{j=1}^{d} \beta_j X_{ij}\right)}{1+exp\left(-\left(\beta_0 + \sum_{j=1}^{d} \beta_j X_{ij}\right)\right)} \right\}$$

$$= log_e \left(exp\left(\beta_0 + \sum_{j=1}^{d} \beta_j X_{ij}\right)\right)$$

$$= \beta_0 + \sum_{j=1}^{d} \beta_j X_{ij} \qquad (4)$$

the function 4 is a generalized linear model (GLM) with d independent variables.

The motivation to the use of logistic model is that it follows the properties of the GLM. Lets define the hypothetical population proportion of cells for which $Y=1$ as $\pi = P(Y=1 \mid X=x)$. Then the theoretical proportion of cells for which $Y=0$ is $1-\pi = P(Y=0 \mid X=x)$. We estimate π by the sample proportions of cells for which $Y=1$. In the GLM context, it is assumed that there exists a set of predictor variables, $X_{11}, X_{12}, \cdots, X_{1d}$, that are related to Y and therefore provides additional information for estimating Y. For mathematical reasons of additivity and multiplicity, logistic model is based on linear model for the log odds in favour of $Y=1$.

$$log_e \frac{\pi_i}{1-\pi_i} = \alpha + \sum_{j=1}^{d} \beta_j X_{ij}$$

thus

$$\pi_i = \sum_{j=0}^{d} \beta_j X_{ij}$$

where $\beta \in \Re^d$ of unknown parameters.

The logistic regression (logit link)

$$g(\pi_i) = log_e \frac{\pi_i}{1-\pi_i} = logit(\pi_i)$$

and

$$g^{-1}(g(\pi_i)) = \pi_i$$

thus the inverse of the logit function in terms of $(X;\beta)$ is given by;

$$g^{-1}(X;\beta) = \pi_i = \frac{1}{1+exp\left(-\left(\beta_0 + \sum_{j=1}^{d} x_{ij}\beta_j\right)\right)}$$

This model can be rewritten as

$$logit(\pi_i) = \sum_{j=0}^{d} \beta_j X_{ij}$$

III. Methodology

3.1 Maximum Likelihood Estimation of the Parameter β

[10] pointed out that estimating the function $P(Y=1 \mid X=x)$ in 1 is equivalent to estimating the function $g(X;\beta)$ 2. Parametric estimation of $g(X;\beta)$ can be found in [15], [21] and [22] among other authors, they used the maximum likelihood estimation method. As they pointed out, one first defines the likelihood function. For the Bernoulli distribution case we have

$$L(Y,X;\beta) = \prod_{i=1}^{n} [P(Y=1 \mid X=x)]^{y_i} [1-P(Y=1 \mid X=x)]^{1-y_i}$$

So, taking the logarithm and upon simplification we have

$$L(Y,X;\beta) = \sum \{(y_i - g(X;\beta) - log_e(1+exp(g(X;\beta)))\} \qquad (5)$$

The regularity conditions requires that the MLEs of β satisfies the usual consistency and asymptotic normality properties [1, 11].

The optimization of the function in 5 with respect to the unknown vector β requires iterative techniques since first derivative is nonlinear in $\hat{\beta}$ and has no simple analytical solution for $\hat{\beta}$ [16].

$$L'(Y,X;\beta) = \sum_{i=1}^{d} y_i x_{ij} - n_i \frac{exp\left(\sum_{i=1}^{d} \beta_j X_{ij}\right)}{1+exp\left(\sum_{i=1}^{d} \beta_j X_{ij}\right)} x_{ij} \qquad (6)$$

$$= \sum_{i=1}^{d} y_i x_{ij} - n_i \pi_i x_{ij} \qquad (7)$$

In matrix form, 7 can be rewritten in the form;

$$L'(Y,X;\beta) = \sum_{i=1}^{n} (y_i - \pi_i)\mathbf{X} \qquad (8)$$

The equation,

$$\pi_i = \frac{exp\left(\sum_{i=1}^{d}\beta_j X_{ij}\right)}{1+exp\left(\sum_{i=1}^{d}\beta_j X_{ij}\right)}$$

is strictly increasing function (monotone) of β_j and approaches 0 as $\beta_j \rightarrow \infty$ and approaches n as $\beta_j \rightarrow \infty$. The second derivative of 6 is strictly negative for all β_j's and as such the solution is a maximum [2, 23].

3.2 Numerical Optimization

The Newton-Rapson method requires that the starting values be sufficiently close to the solution to ensure convergence. Under this condition the Newton-Raphson iteration converge quadratically to at least a local optimum. When the method is used to the problem of maximizing the likelihood function, it produces a squence of values $\theta^{(0)}, \theta^{(1)}, \cdots, \theta^{(\theta)}$ that under ideal conditions converge to the MLEs $\hat{\theta}_{mle}$.

the motivation to the use of the method is that this approximation is valid provided the unknown parameter β^{j+1} is in the neighbourhood of β^j. Since $L(Y,X;\beta)$ corresponds to the objective function to be minimized, $L'(Y,X;\beta)$ represents the gradient vector, the vector of first order partial derivative and $I(\theta)$ to the negative of the Hessian matrix $H(\theta)$ which is a matrix of the second order derivative of the objective function $L''(Y,X;\beta)$. Then the Hessian matrix is used to determine whether the minimum of the objective function $L(Y,X;\beta)$ is achieved by the solution $\hat{\theta}$ for the equation $L'(Y,X;\beta)=0$, that is, whether $\hat{\theta}$ is a stationary point of $L(Y,X;\beta)$. If this is the case the $\hat{\theta}$ is the maximum likelihood estimate of the matrix of θ the iterative formula for finding a maximum or minimum of a function $f(x)$ is given by ;

$$X^{(j+1)} = X^{(j)} - H_i^{-1} l'(\theta)$$

where H_i is the Hessian $f''(X_i^j)$ and $l'(\theta)$ is the gradient vector, $f'(x)$ of $f(x)$ at the i^{th} iteration.

Then the i^{th} iteration is given by;

$$\hat{\theta}^{(j+1)} = \hat{\theta}^{(j)} - \left(H(\hat{\theta})^j\right)^{-1} l'(\theta) \quad (9)$$

In otherwords,

$$\hat{\theta}^{(j+1)} = \hat{\theta}^{(j)} - \frac{l'(\theta)}{l''(\theta)}$$

which is the iterative generator.

But from 9

$$L'(Y,X;\beta) = \sum_{i=1}^{d} y_i x_{ij} - n_i \pi_i x_{ij}$$

In matrix form;

$$L'(Y,X;\beta) = \sum_{i=1}^{n} (y_i - \pi_i)\mathbf{X}$$

and the negative of the second derivative;

$$I(\beta) = \frac{\partial^2 L(Y,X;\beta)}{\partial\beta\partial\beta'} = \sum_{i=1}^{n} \pi_i (1-\pi_i)\mathbf{X}'\mathbf{X}$$

The matrix $I(\beta)$ is a $(p+1)\times(p+1)$ matrix. The matrix plays a key role in the estimation procedure and yields the logit estimates obtained by inverting the Hessian (or expected Hessian) matrix or the information matrix. Then the Newton-Raphson iterative solution of the system of equations can be used to obtain the solution of $\beta' s$. At the i^{th} iteration, estimates are obtained as;

$$\hat{\beta}^{(i+1)} = \hat{\beta}^{(i)} - \left[I\left(\hat{\beta}^{i}\right)\right]^{-1} L'\left(Y, X; \hat{\beta}^{i}\right) \qquad (10)$$

where the least square estimates of the $\beta' s$ are used as initial estimates.

Continue applying Equation 10 until there is essentially no change between the elements of β from one iteration to the next. At that point, the maximum likelihood estimates are said to converge.

IV. Simulation study

4.1 Checking consistency of the maximum likelihood Estimators

Nonlinear system of equations arise commonly in statistic. In some cases, there will be a naturally associated scalar function of parameters which can be optimized to obtain parameter estimates. The MLE cannot be written in closed form expression, thus substantialy complicating the task of evaluating the characteristic of its (finite sample) distribution, whether the variables are random or not. Maximum likelihood estimator simulation for large samples are carried out using the Monte-Carlo simulation method. The simulations of the study involves the regressor variables which are fixed and for each model parameter, n-simulation binomial data set are generated for each of the regressor variable $x_1, x_2, \cdots, x_n$. We consider the complete model to be simulated as;

$$\begin{aligned} y_i &= g(X;\beta) + e_i \\ &= 1 if\ X_i\beta + e \geq a \\ &= 0 if\ X_i\beta + e < a \end{aligned}$$

where y_i is the dependent variable to incorporate the effects of the independent variables. The row vector X_i represents the i^{th} observations on all predictor variables.

The basic model can be structured as

$$\begin{aligned} \pi_i &= Pr(y_i = 1 \mid x_i) \\ 1 - \pi_i &= Pr(y_i = 0 \mid x_i) \end{aligned}$$

For the logit model;

$$\pi_i = \frac{exp(x'\beta)}{1 + exp(x'\beta)}$$

which is the cdf of the logistic distribution.

For each generated data set, the mle for $\hat{\beta}$ is computed and saved. This procedure is repeated for $n = 200, 300, 500,$ and 700 at each of the regressor levels.

The following table gives the results of the simulation study for different sample sizes.

Table 1: Estimated-parameter values and their standard errors using the regression model for different sample sizes

β	$n = 200$		$n = 300$		$n = 500$		$n = 700$	
	Estimates	SE	Estimates	SE	Estimates	SE	Estimates	SE
β_0	-42.356	472.855	-24.268	12.830	-22.872	3.583	-22.497	2.947
β_1	6.237	73.156	3.425	2.173	3.236	0.436	3.177	0.357
β_2	0.310	4.085	0.136	0.079	0.129	0.057	0.127	0.047
β_3	0.039	0.716	0.036	0.024	0.034	0.018	0.033	0.015
β_4	1.501	2.204	0.983	0.29	0.937	0.206	0.920	0.169

As seen in the table 0, as the sample size increases from $n = 200$ to $n = 700$ the estimated values of the parameters are very close to the true values $\beta_0, \beta_1, \beta_2, \beta_3, \text{and} \beta_4$ and the standard deviations of the estimates are noticeably smaller. This indicates that this simulation study performs well in showing the consistency of the maximum likelihood estimators for parameters of the logistic model.

4.2 Regularity conditions of the asymptotic normality of a Binomial Response model

[9] present regularity conditions for a very general class of generalized linear models. In this section, we explain the regularity conditions under the Binomial response model and then we apply Theorem **1** to show the asymptotic properties of ML estimators for the Binomial response model.

(C1): The pdf $g(X;\beta)$ is distinct, that is $\beta \neq \beta'$ implying that $g(X;\beta) \neq g(X;\beta')$, thus the model is identifiable.

The proof of this assumption has been well documented by [23]

(C2): The pdf have common support for all β, the true parameter vector is in the interior of this space.

This condition holds if the domain (support) of X is a closed set [18].

[18] noted that the restriction that true parameter vector in the interior excludes some cases where consistent and asymptotically normal (CAN) breaks down. This is not a restrictive assumption in most application, but it is for some.

(C3): The response model is measurable in x, and for almost all x is continous in the parameters. The standard models such as the probit, logit and the linear probability model are all continous in their argument and in x, so that this assumption holds.

(C4): The model satisfies a global identification (that is it guarantees that there is at most one global maxima, see [18].

The proof of this assumption has been discussed well by [23]. The concavity of the log-likelihood of an observation for the logit guarantees global identification, provided only that the $x's$ are not linearly independent.

(C5): The assumption states that the model log likelihood is twice or three times differentiable, this is true provided the parameters do not give observations on the boundary in the linear or log linear models where probabilities are zero or one. [8] shows that these conditions are specifically satisfied for the binomial model.

(C6): The log likelihood and its derivative have bounds independent of the parameters in some neighbourhood of the true parameter values. The first derivative have the Lipschitz property in the neighbourhood. This property is satisfied by the logistic model since it is continously differentiable (McFadden,1999).

(C7): The pdf $g(X;\beta)$ is three times differentiable as a function of β. Further, for all $\beta \in \Omega$, there exists a constant c and a function $M(x)$ such that for all $\beta_0 - c < \beta < \beta_0 + c$ and all x in the support of X.

$$\left| \frac{\partial^3}{\partial \beta^3} log\, g(X;\beta) \right| \leq M(x)$$

with

$$E_{\beta_0}[M(X)] < \infty$$

for all $\beta_0 - c < \beta < \beta_0 + c$ and all x in the support of X. The proof of this assumption has been done by many authors like [2, 23]. This implies that the information matrix, equal to the expectation of the outer product of the score of an observation is non-singular at the true parameter.

The conditions $(C1), \cdots, (C7)$ may seem restrictive at first, but are met for a wide range of link functions. The results guarantee that the MLE estimates of β is essentially carried out by linearizing the first order condition for the estimator using a Taylor's expansion. Since the binomial model satisfies the above conditions, then following theorem holds for the parameter $\bar{\beta}$.

1 *Let* $x_1, x_2, x_3, \cdots, x_n$ *be iid each with a density* $g(x;\beta)$. *Then, with probability tending to 1 as*

$n \to \infty$, *there exists solutions* $\hat{\beta}_n = \hat{\beta}(x_1, \cdots, x_n)$ *of the likelihood equations.*

$$\frac{\partial}{\partial \beta_j}\left[g(x_1;\beta), \cdots g(x_n;\beta) \right] = 0, \qquad j = 1,2,\cdots,n$$

or equivalently

$$\frac{\partial}{\partial \beta_j}\left[\log L(\beta) \right] = 0, \qquad j = 1,2,\cdots,d$$

such that

(a) $\hat{\beta}_{jn}$ is consistent for estimating β_j.

(b) $\sqrt{n}(\hat{\beta}_n - \beta)$ is asymptotically normal with mean (vector) zero and covariance matrix $[L(\beta)^{-1}]$, and

(c) $\hat{\beta}_{jn}$ is asymptotically efficient in the sense that

$$\sqrt{n}(\hat{\beta}_{jn} - \beta_j) \underline{L} N\left(0, [I(\beta)]_{jj}^{-1}\right)$$

4.3 Normality of the ML estimators

Under some assumptions that allows among several analytical properties, the use of the delta method, the central limit theorem holds. We conducted a simulation study via the freeware package R. We show how the properties of an estimator are affected by changing conditions such as its sample size and the value of the underlying parameters. Employing it in practice, we illustrate the large sample behavior of the estimated parameters $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \text{and} \hat{\beta}_4)'$ and also look at the sensitivity of the QQ-plots using the Shapiro-Wilks and the Kolmogorov-Smornov test, we show that;

$$\sqrt{N}\left(\hat{\beta}_{mle} - \beta\right) \to N\left(0, \frac{1}{I(\beta_{mle})}\right) \qquad (11)$$

where

$$I(\beta) = -E_\beta \begin{pmatrix} \frac{\partial^2 \log l}{\partial \beta_0^2} & \frac{\partial \log l}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 \log l}{\partial \beta_0 \partial \beta_1} & \frac{\partial \log l}{\partial \beta_0 \partial \beta_2} & \frac{\partial \log l}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 \log l}{\partial \beta_1 \partial \beta_0} & \frac{\partial \log l}{\partial \beta_1^2} & \frac{\partial^2 \log l}{\partial \beta_1 \partial \beta_2} & \frac{\partial \log l}{\partial \beta_1 \partial \beta_3} & \frac{\partial \log l}{\partial \beta_1 \partial \beta_4} \\ \frac{\partial^2 \log l}{\partial \beta_2 \partial \beta_0} & \frac{\partial \log l}{\partial \beta_2 \partial \beta_1} & \frac{\partial \log l}{\partial \beta_2^2} & \frac{\partial \log l}{\partial \beta_2 \partial \beta_3} & \frac{\partial \log l}{\partial \beta_2 \partial \beta_4} \\ \frac{\partial^2 \log l}{\partial \beta_3 \partial \beta_0} & \frac{\partial \log l}{\partial \beta_3 \partial \beta_1} & \frac{\partial \log l}{\partial \beta_3 \partial \beta_2} & \frac{\partial \log l}{\partial \beta_3^2} & \frac{\partial \log l}{\partial \beta_3 \partial \beta_4} \\ \frac{\partial^2 \log l}{\partial \beta_4 \partial \beta_0} & \frac{\partial^2 \log l}{\partial \beta_4 \partial \beta_1} & \frac{\partial^2 \log l}{\partial \beta_4 \partial \beta_2} & \frac{\partial \log l}{\partial \beta_4 \partial \beta_3} & \frac{\partial \log l}{\partial \beta_4^2} \end{pmatrix}$$

For different sample sizes of $n = 500, 700, 1000, 1500 \quad , \text{and} 2000$, we calculate the Equation 11 and repeat it 5,000 times. The results are presented in the Figures 1, 2, 3, 4 and 5, through the quantile-quantile normal plot for $\hat{\beta}$.

A quantile-quantile normal graph, plots the quantiles of the data set against the theoretical quantiles of the standard normal distribution. If the data set appears to be a sample from a normal population, then the points will fall roughly along the line. The computation results indicates that the distribution of parameters approximates normal distribution as sample size, n increases.

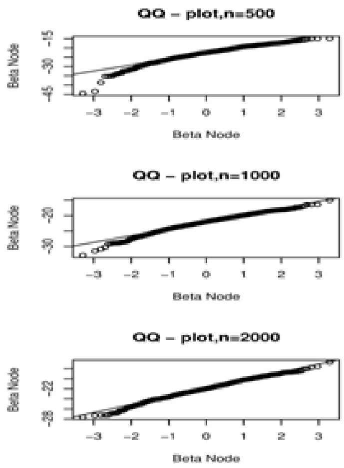

Figure 2: Monte Carlo Simulation of finite sample behaviour for normality of the parameter $\hat{\beta}_0$

Table 2: Test for Nomality β_0

	Kolmogorov-Smirnov test		Shapiro-Wilks test	
Sample size(n)	Test statistic (D)	P-value	Test statistic (D)	P-value
500	0.0501	0.09864	0.9948	0.0016
700	0.0664	0.0100	0.9961	0.0119
1000	0.0389	0.3246	0.9964	0.01997
1500	0.0325	0.5600	0.9987	0.3417
2000	0.0323	0.5567	0.9986	0.0462

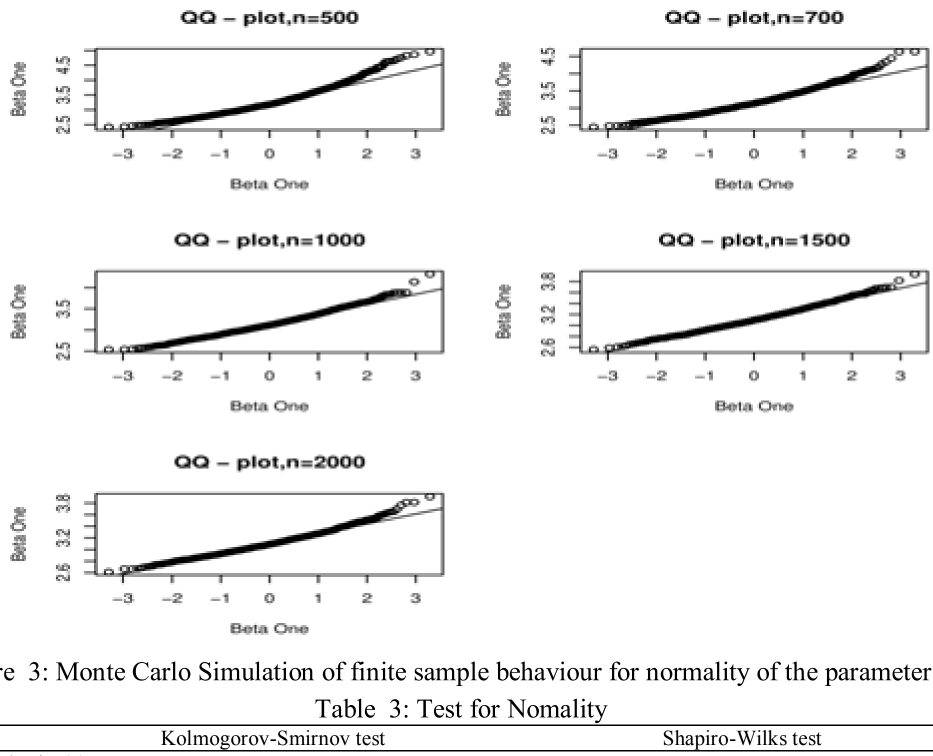

Figure 3: Monte Carlo Simulation of finite sample behaviour for normality of the parameter $\hat{\beta}_1$

Table 3: Test for Nomality

	Kolmogorov-Smirnov test		Shapiro-Wilks test	
Sample size(n)	Test statistic (D)	P-value	Test statistic (D)	P-value
500	0.0600	0.0015	0.0600	0.0015
700	0.0584	0.0037	0.0654	0.0004
1000	0.0493	0.0156	0.0493	0.0156
1500	0.0431	0.0491	0.0431	0.0491
2000	0.0312	0.2846	0.0312	0.2846

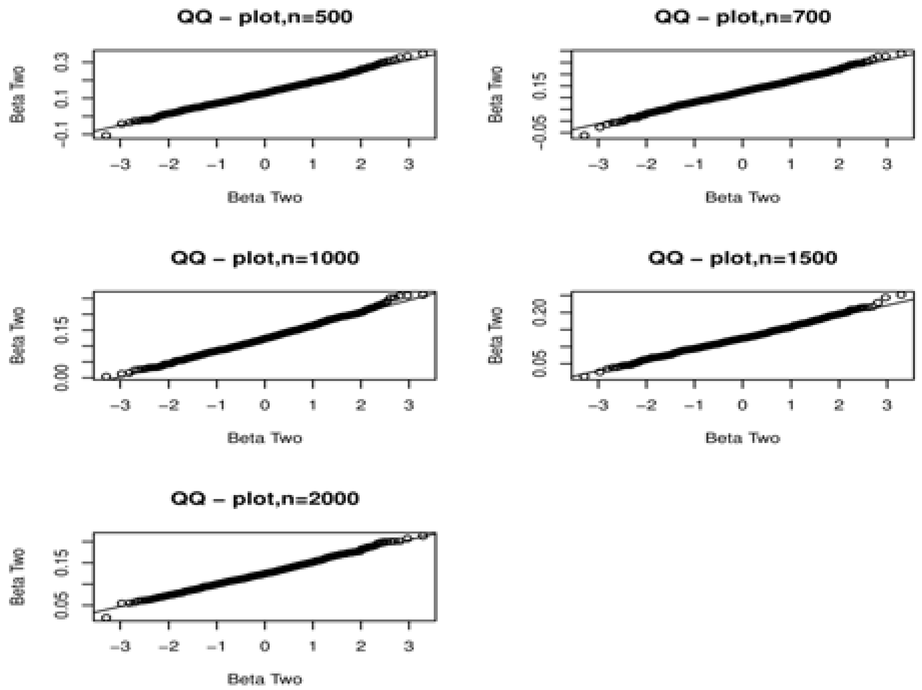

Figure 4: Monte Carlo Simulation of finite sample behaviour for normality of the parameter $\hat{\beta}_2$

Table 4: Test for Nomality

	Kolmogorov-Smirnov test		Shapiro-Wilks test	
Sample size(n)	Test statistic (D)	P-value	Test statistic (D)	P-value
500	0.0309	0.2948	0.9998	0.0899
700	0.0346	0.1831	0.9970	0.1934
1000	0.0326	0.2378	0.9961	0.05678
1500	0.0295	0.3457	0.9974	0.0403
2000	0.0291	0.3661	0.9995	0.1101

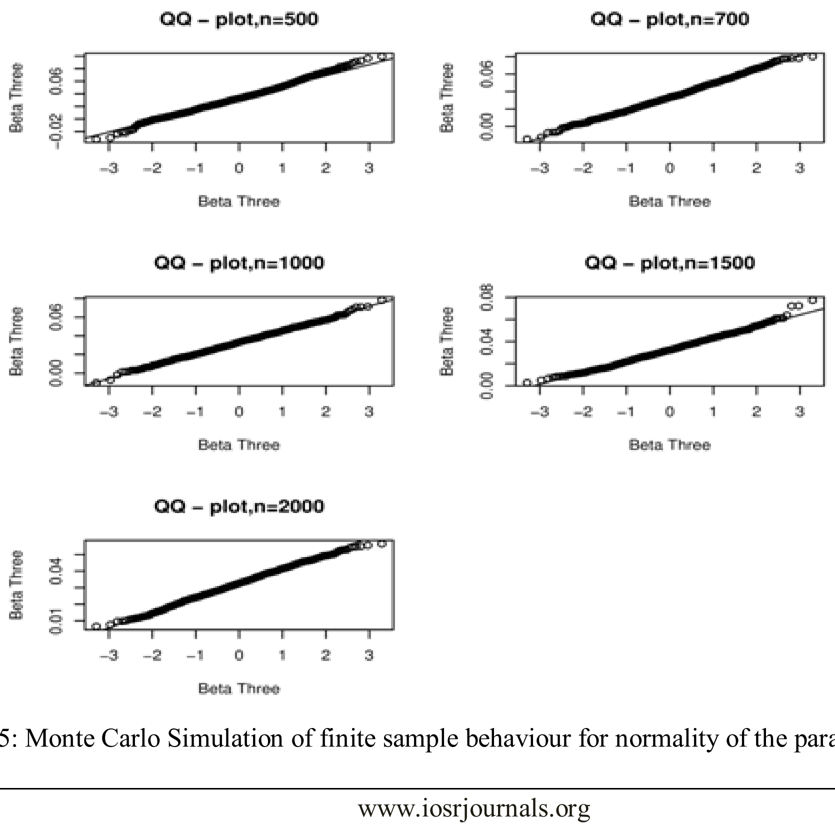

Figure 5: Monte Carlo Simulation of finite sample behaviour for normality of the parameter $\hat{\beta}_3$

Table 5: Test for Nomality

	Kolmogorov-Smirnov test		Shapiro-Wilks test	
Sample size(n)	Test statistic (D)	P-value	Test statistic (D)	P-value
500	0.0315	0.2731	0.9930	0.0001
700	0.0287	0.3825	0.9952	0.0029
1000	0.0167	0.5498	0.9969	0.0471
1500	0.0122	0.8700	0.9945	0.0707
2000	0.0096	0.8374	0.9988	0.7674

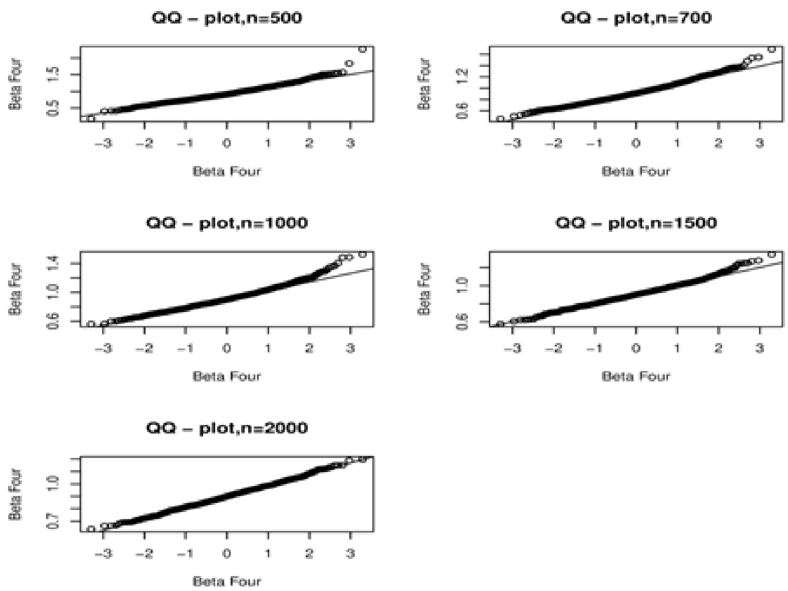

Figure 6: Monte Carlo Simulation of finite sample behaviour for normality of the parameter $\hat{\beta}_4$

Table 6: Test for Nomality

	Kolmogorov-Smirnov test		Shapiro-Wilks test	
Sample size(n)	Test statistic (D)	P-value	Test statistic (D)	P-value
500	0.0426	0.0529	0.9958	0.0084
700	0.0363	0.1431	0.9916	.01843
1000	0.0459	0.2952	0.9968	.04791
1500	0.0225	0.6946	0.9973	0.09807
2000	0.0187	0.9001	0.9995	0.9980

V. Conclusion

The study shows that the asymptotic properies of the maximum likelihood estimates of the logistic regression model can be obtained by some transformation of the regularity conditions of the linear regression model. The simulation studies done show that there is consistency in the parameter estimates, where fixed values of regression parameters are used, this shows that simulated estimates converge well to the fixed values as the sample size approaches infinity. The finite behaviour of consistency is upheld.

On the otherhand, simulated result on the normality were taken using the Q-Q-plots and using the the Kolmogorov-Smirnov and Shapiro-Wilks test. The analysis shows that the parameters are normally distributed, this can be checked on the decrease of the statistic values on both tests and also from tables 1, 2, 3, 4 and 5, we see that we fail to reject the null hypothesis at $\alpha = 5\%$ as the sample size increases and conclude that the samples are taken from the normal distribution.

References

[1] T. Amemiya. Advanced Econometric. Havard University Press, Cambridge, 1985.
[2] M. Beer. Asymptotic properies of the maximum likelihood estimator in [1] dichotomodi logistic regression model. 2001.
[3] David Collett. Modelling Binary Data. Chapman & Hall/CRC, New York, USA, second edition, (2002).
[4] Jerome Cornfield. Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: A discriminant function analysis. In Federation Proceedings, volume 21, page 58, 1962.
[5] D. R. Sir Cox and E. J. Snell. Analysis of Binary Data. London.Chapman & Hall, 1989.
[6] J. S. Cramer. Logit Models from Economics and Other Fields. Cambridge University Press, 2003.
[7] K. S. Crump, H. A. Guess, and K. I. Deal. Confidence intervals and test of hypothesis concerning dose response relations inferense. 1977.

[8] R. C. Deutsch. Phd thesis. 2007.
[9] I. Fahrmeir and H. Kaufmann. Consistency and asymptotic normality of maximum likelihood estimator in generalized linear models. 1985.
[10] J. Fan, M. Farmen, and J. Gijbels. Local maximum likelihood estimator and inference. 1998.
[11] C. Gourienx and A. Monfort. Asymptotic properties of the maximum likelihood estimation in dichotomous logit models. Journal of Econometric 17, 83–97, 1981.
[12] H. A. Guess and K. S. Crump. Maximum likelihood estimation of dose response models subject absolutely monotonic constraints. 1978.
[13] D. W Hosmer and S. Lemeshow. Applied Logistic Regression. New York, Wiley, 1989.
[14] Trina Hosmer, David Hosmer, and Lloyd Fisher. A comparison of the maximum likelihood and discriminant function estimators of the coefficients of the logistic regression model for mixed continuous and discrete variables. Communications in Statistics-Simulation and Computation, 12(1):23–43, 1983.
[15] D. N. Joanes. Reject inference applied to logistic regression for credit scoring. 1993.
[16] G.S Maddala. Limited dependent variables and quantitative variables. 1983.
[17] P. McCullagh and J. A. Nelder. Generalized linear models. 1989.
[18] L. McFadden. Conditional logit analysis of qualitative choice behaviour. 1974.
[19] L. Nordberg. Asymptotic normality of maximum likelihood estimators based on independent unequal distributed observations in exponential family models. 1980.
[20] Leandro Pardo. Statistical inference based on divergence measures. CRC Press, 2005.
[21] R. Pastor-Barriuso, E. Guallar, and J. Coresh. Use of two-segmented logistic regression to estimate change of point in epidemological studies. 1998.
[22] R. Pastor-Barriuso and E. Guallar. Transition model for change point estimation on logistic regression. 2003.
[23] M. Rashid and N. Shifa. Consistency of the maximum likelihood estimator in logistic regression model:a different approach. 2009.
[24] R.Christensen. Log-Linear Models and Logistic Regression. Springer-Verlag Inc., New York, USA, 1997.
[25] J Wilde. A note on gmm estimator of probit models with endogenous regression . 2008.

Appendix 8: Article II

Multidimensional Analysis of the Determinants of Poverty Indicators in the Lake Victoria Basin(Kenya).

IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 11, Issue 3 Ver. V (May - Jun. 2015), PP 73-87
www.iosrjournals.org

Multidimensional Analysis of the Determinants of Poverty Indicators in the Lake Victoria Basin(Kenya)

[1]Anthony Ngunyi, [2]Peter. N. Mwita, [2]Romanus O. Odhiambo, [3]Verdiana. G. Masanja

[1,]Dedan Kimathi University of Technology (Kenya)P.O BOX 657-10100, Nyeri
[2,] Jomo Kenyatta University of Agriculture & Technology (Kenya) P.O BOX 62000-00200, Nairobi
[3,] University of Rwanda(Head Quarters) P.O Box 4285, Kigali, Rwanda
[3,]Affiliated to University of Dar es Salaam, P.O.Box 35062 ,Dar es Salaam,Tanzania

Abstract: *The study main objective is to examine the multidimensional aspects of poverty in one Kenya's culturally diverse region of the Lake Victoria basin. The analysis using data collected by IUCEA researchers in 2007 and also the 2009 census on households in Kenya. This study investigates statistical models based on factors that characterize the demographic characteristic of individuals, in determining the predictors of poverty for better policy formulation.. The research findings indicate that poverty measures do overlap to capture a percentage of the sample as poor. The analysis shows that education, gender (being male), marital status, assets (livestock, water sources, and wall materials) and age of the head of the family have statistically positive effects on the likelihood of an individual falling into poverty.*

Keywords: *Poverty, Demography, Augmented, Logistic, Assets*

I. Background Information

According to the World Bank,(2010).[1], "poverty is pronounced deprivation in well-being." This of course begs the questions of what is meant by well-being and of what is the reference point against which to measure deprivation.

The objective of the study was to look at the different factors that influence poverty in a household, and the policy formulation that can be put in place in order to achieve bettering living standard for the members of the household. This study based its results on a multidisplinary aspects on the fact that many studies on poverty in Kenya have been on regressing well known determinants even though other factors may be able to give an informative and simple to interpret facts on poverty levels in the region.

One approach is to think of well-being as the command over commodities in general, so people are better off if they have a greater command over resources. The main focus is on whether households or individuals have enough resources to meet their needs, see, S. Pudney(1999) [2]. Typically, poverty is then measured by comparing individuals' income or consumption with some defined threshold below which they are considered to be poor. This is the most conventional view-poverty is seen largely in monetary terms-and is the starting point for most analysis of poverty.

A second approach to well-being (and hence poverty) is to ask whether people are able to obtain a specific type of consumption: Do they have enough food? Or shelter? Or health care? Or education? As cited in Ravallion and Bidani (1994); Kakwani (1990),[3,4]. In this view the analyst goes beyond the more traditional monetary measures of poverty: Nutritional poverty might be measured by examining whether children are stunted or wasted; and educational poverty might be measured by asking whether people are literate or how much formal schooling they have received, well articulated in Lipton and Ravallion (1995),[5].

Perhaps the broadest approach to well-being is the one articulated by Sen (1999), [6], who argues that well-being comes from a capability to function in society. Thus, poverty arises when people lack key capabilities, and so have inadequate incomes or education, or poor health, or insecurity, or low self-confidence, or a sense of powerlessness, or the absence of rights such as freedom of speech. Viewed in this way, poverty is a multidimensional phenomenon and less amenable to simple solutions. For instance, while higher average incomes will certainly help reduce poverty, these may need to be accompanied by measures to empower the poor, or insure them against risks, or to address specific weaknesses such as inadequate availability of schools or a corrupt health service (Datt and Jolliffe, 2005). [7].

WHO (2000),[8] noted that poverty is related to, but distinct from, inequality and vulnerability. Inequality focuses on the distribution of attributes, such as income or consumption, across the whole population. In the context of poverty analysis, inequality requires examination if one believes that the welfare of individuals depends on their economic position relative to others in society. Vulnerability is defined as the risk of falling into poverty in the future, even if the person is not necessarily poor now; it is often associated with the effects of "shocks" such as a drought, a drop in farm prices, or a financial crisis. Vulnerability is a key dimension of well-being since it affects individuals' behavior in terms of investment, production patterns, and coping strategies, and in terms of the perceptions of their own situations.

According to the last Country Briefs, an estimated 3.8 million people in rural areas are between highly to extremely food insecure. Food and Agriculture Organization (FAO)/ Global Information and Early Warning System on Food and Agriculture (GIEWS) and Famine Early Warning System (FEWSNET) agree that, in the short term, Kenya is a hunger-prone country, while WFP and IFPRI assess the long-term situation as alarming and hunger as moderately high.. There is a long history of periodic shortfalls in food supply in Kenya. Shortfalls occur all over the country or in parts of the country, and sometimes for two years in a row. In times of unfavorable weather, even the provinces normally characterized by a maize surplus (such as the Rift Valley) or marginally self-sufficient provinces (such as Western and Nyanza) may enter a maize deficit situation. In addition, in areas characterized by chronic deficits (such as the Coast and Eastern and North Eastern provinces) the situation becomes acute. In many districts in these areas, emergency relief becomes necessary.

The highest poverty rate was found among people living in households headed by farmers 46 percent (KNBS, 2007a), [11]. By contrast, households headed by someone working in the government are least likely to be poor; in these occupations the poverty rate was 20 percent (1993–94). This would suggest that policies that aim to reduce poverty through enhancing income-generating capabilities should be targeted towards the agricultural sector.

The relationship between poverty and education is particularly important because of the key role played by education in raising economic growth and reducing poverty. The better educated have higher incomes and thus are much less likely to be poor. Kenyans living in households with an uneducated household head are more likely to be poor, with a poverty rate of 47 percent in 2014 national poverty atlas.. With higher levels of education, the likelihood of being poor falls considerably. Raising education attainment is clearly a high priority to improve living standards and reduce poverty.

The relationship between gender and poverty may also indicate another targeting strategy for poverty reduction. In Tanzania, about 35 percent of the population lives in households headed by women. Perhaps surprisingly, the 2007 data show that the poverty rate was slightly lower among female-headed households (48 percent) than among male-headed households (52 percent). In this case, targeting interventions based on the gender of the head of household would not help to distinguish the poor from the non-poor, Mark Schreiner, [13].

II. Literature Review

Poverty is a worldwide concern. Although there is a global concern towards poverty reduction, there is a little agreement on a single definition and measurement of poverty (Kotler et al., 2006; Laderchi et al., 2003), [14, 15]. According to Kotler et al., (2006),[14] and Laderchi et al.(2003),[15], the problem of arriving at one single definition of poverty has been compounded by a number of factors. Poverty affects heterogeneous groups such that the concept of poverty is relative depending on different interest groups and individuals experiencing it (Kotler et al., 2006, Rank, 2004), [14, 16]. The difficulty surrounding the definition and measurement of poverty has often led poverty researchers and policy makers to relate poverty to the concepts of impoverishment, deprivation, the disadvantaged, inequality, the underprivileged and the needy.

Many researchers have authored many articles on the issue of poverty worldwide. The exception being the absolute poverty measures for the developing world by Chen and Ravallion (2007) [1], which serve to provide the latest evidence for an African exceptionalism that dominates the development needs of today.

All developing country regions have shown marked improvement in key indicators of poverty,health, economy, and food, except for sub-Saharan Africa. For poverty, the global number of people living below the extreme poverty line of $1 per day decreased between 1981 and 2004 from 1,470 million to 969 million. The percentage of extremely poor fell from 40% to 18%. However, in sub- Saharan Africa, the numbers almost doubled from 168 million to 298 million, and the percentage stayed almost constant from 42% to 41% , Chen S, Ravallion M (2007) [35].

For health, the life expectancy at birth in sub-Saharan Africa peaked in 1990 at 50 years but has since declined to 46 years, while steadily rising in all developing country regions to an average of 65 years, Jamison D.T, (2006),[36]. Over the period 1960–2000, sub-Saharan Africa's per capita measure of annual economic growth (gross domestic product) was a mere 0.1%, whereas other developing country regions experienced accelerated growth averaging 3.6%, Collier P (2007), [37]. Food production per capita grew by 2.3% per year between 1980 and 2000 in Asia, grew by 0.9% in Latin America, and declined by 0.01% in tropical Africa see, Dasgupta .P et al (2004),[38].

There are basically two approaches in modelling determinants of poverty. The first approach[5] is the employment of consumption expenditure per adult equivalent and regress it against potential explanatory variables (Geda et al, 2001). Using this approach Arneberg and Pederson (2001) report that household characteristics and education are the main factors which affect living standard in Eritrea. However, they treat education as a linear and continuous variable. Moreover they find out that transfer payment from relatives abroad is a significant contributor to the welfare of a society. From their analysis they conclude that education is the most important factor for the way out of poverty. However, their approach suffers from the common problems of consumption as being indicator of welfare and the assumption that consumption of the poor and non poor are both determined by the same process (Okwi, 1999). The second approach is to directly model poverty by employing a discrete choice model.

The practice of discrete choice models in the analysis of determinants of poverty has been popular approach[6] (for instance, Fafack(2002) for Burkian'faso, Kabubuo-Mariara (2002) for Kenya; Amuedo_Dorantes(2004) for Chile; Grootaert(1997) for Cote D'voire; Geda et al (2001) for Kenya; Charlette-Gueard and Mesple-Somps (2001) for Cote d`voire , Goaed and Ghazouani (2001) for Tunisia; Roubaud and Razafindrakoto ,2003). The analysis then proceeds by employing binary logit or probit model to estimate the probability of a household being poor conditional up on some characteristics. In some cases also the households are divided into three categories: absolute poor, poor and non poor and then employ ordered logit or ordered logit model to identify the factors which affect the probability a household being poor conditional up on set of characteristics. In this study we apply the dicrete choice model as discussed by many researchers in kenya but also look at the augmented model proposed by Datt. G and Jolliffe .D. (2005),[34]

Common indices developed by the United Nations Development Programme are the human development index composed of three measures of development (per capita gross domestic product, life expectancy, and literacy) or the human poverty index composed of measures of deprivation in the development indices (child and young adult mortality, illiteracy, and lack of water and sanitation) United Nations Development Programme (2006),[37].In the study of the lake Victoria basin ,we look at the aspects of the asset component as a measure of poverty and articulate the best policy measures that can be taken into consideration to reduce poverty in the area.

Poverty studies in Kenya have focused on a discussion of inequality and welfare based on limited house level data (Arne, 1981; Hazlewood, 1981; House and Killick, 1981) [17,18,19]. One recent comprehensive study on the subject is that of Geda et al. (2001), [20], which deals with measurement, profile and determinants of poverty. The study employs a household welfare function, approximated by household expenditure per adult equivalent. The authors runs two categories of regression, using overall expenditures and food expenditures as dependent variables. In each of the two cases, three equations are estimated which differ by type of dependent variable. These dependent variables are: total household expenditure, total household expenditure gap (the difference between the absolute poverty line and the actual expenditure) and the square of the latter. A similar set of dependent variables is used for food expenditure, with the explanatory variables being identical in all cases.

Geda et al. (2001), [20] , justified their choice of this approach (compared to a logit/probit model) as follows; First, the two approaches (discrete and continuous choice-based regressions) yield basically similar results; also the expenditure as a binary variable has certain inherent weakness. One obvious weakness is that, unlike the logit/probit model, the level of the regression about poverty. Second, the major assumption of the welfare function approach is that consumption expenditure are negatively associated with absolute poverty at all expenditure levels. Thus factors that increase consumption expenditure reduce poverty. However, this basic assumption needs to be taken cautiously. For instance though increasing welfare, raising the level of consumption expenditure of households that are already above the poverty line does not affect the poverty level (for example measures by the headcount ratio). Notwithstanding such weakness, the approach is widely used

Geda et al. (2001),[20] identified the following as important determinants of poverty: unobserved region-specific factors, mean age, size of household, place of residence (rural versus urban), level of schooling, livestock holding and sanitary conditions. The importance of these variables does not change whether the total expenditure, the expenditure gap or the square of the gap is taken as the dependant variable. The only noticeable change is that the sizes of the estimated coefficients are enormously reduced in the expenditure gap and in the square of the expenditure gap specifications. Moreover, except for the minor changes in the relative importance of some of the variable, the pattern of coefficient again fundamentally remains unchanged when the regressions are run with food expenditures as dependant variable.

Another recent study on the determinant of poverty is Oyugi (2000),[21], which is an extension to earlier work by Greer and Thorbecke (1986b,a).[22,23]. The later study used household calorie consumption as the dependant variable and a limited number of household characteristics as explanatory variables. An important aspect of Oyugi's study is that it analyse poverty both at micro (household) and meso (district) level, with the meso level analysis being the innovative component of the study. The explanatory variable (household characteristics) include: holding area livestock unit, the proportion of household members able to read and write, household size, sector of economic activity (agriculture, manufacturing/industrial). The results of the probit analysis show that all variable used are important determinants of poverty in rural areas and at the national level, but that there are important exceptions for urban areas.

In the probit model, however, in the order of importance the key determinants of poverty are: being able to read and write, employment in off-farm activities, being engaged in agriculture, having a side-business in the service sector, source of water and household size. Region of residence appears to be equally important in determining poverty status in the two approaches. Although the two approaches did not employ the same explanatory variables, this comparison points to the possibility of arriving at different policy conclusions from the two approaches Oyugi (2000),[21].

III. Methodology

3.1 Area of Study

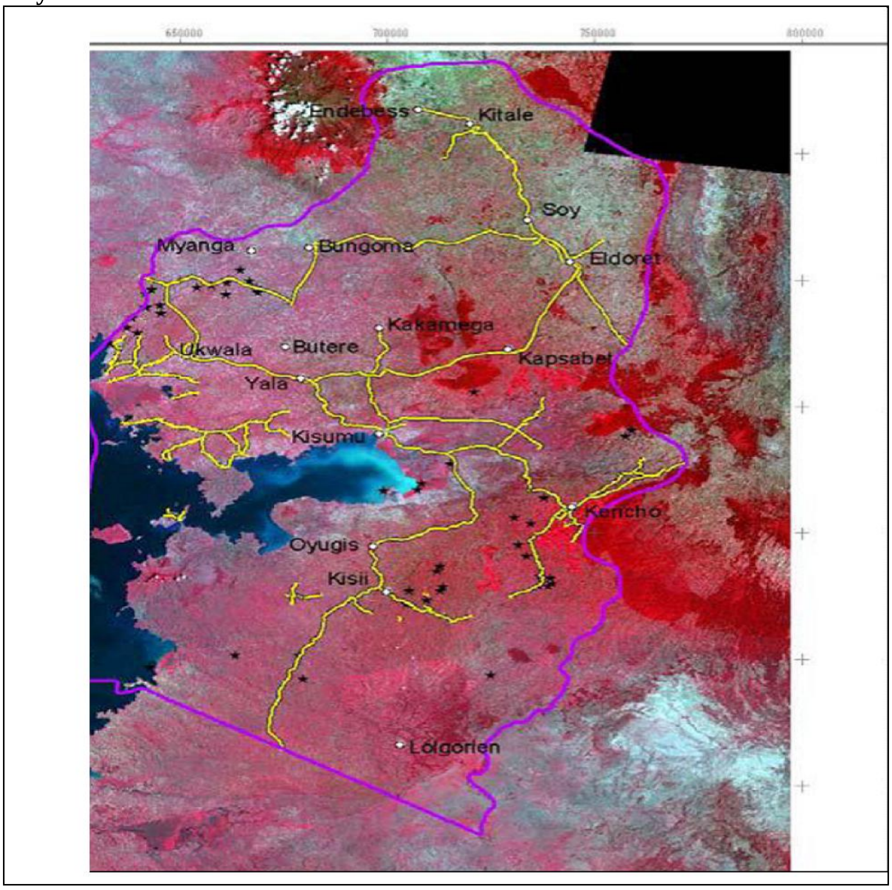

Figure 1: The Lake Victoria Basin on the Kenyan side.

The study site constitutes three districts of the Lake Victoria basin. Some of the raw data on Household Demography, collected during the first year work of the project entitled "Mathematical Techniques for Food Crops Balance Sheet and Food Security Indicators in Lake Victoria Water Shed", see [24] is used in this study. The random sampling approach was employed to select the study areas and sample respondents in which the subjects selected were supposed to meet the study needs. A total of 24 households in each of the three districts (Kuria, Siaya and Kisumu) in Kenya were surveyed using structured questionnaire, interview sessions, focus group discussion and observation. A list of household heads (which is the sampling frame from which a probability sample is selected) were supplied by respective sub-location administrations. These lists were each used to select 24 households from each sub-location by employing simple random sampling technique. This method of sample selection is free form bias; it has given every household head in each sub-location a chance of being included in the sample for this study.

The study also makes use of data obtained from the 2009 Population and Housing Census conducted by the Kenya National Bureau of Statistics. The survey questionnaire collected information on household and demographic characteristics, education, assets, employment, income, and expenditures and assets in the households. The questionnaire included information on household members and was administered to all households in the country, with the exception of North Eastern Province. Although the census did not collect information on income and expenditures, it provides information on a number of characteristics that have been shown to be strong correlates of poverty. . Such characteristics include assets, education and the household size.

3.2 Specification of the Regression Model

When poverty is defined as the current consumption deficit, a household is categorized as poor if the value of per capita consumption of its members is lower than the poverty line. Therefore, it is logical to search for poverty predictors based on variables that correlate with per capita household consumption. These variables can be obtained by estimating a model of consumption correlates, where the left-hand side is per capita consumption and the right-hand side is a set of variables that is thought of correlating with household consumption. Different from determinants model, in correlates model the endogeneity of the right-hand side variables is not a concern,see Datt and Jolliffe, 2005). [37].

Once the set of the right-hand side variables has been determined, a stepwise regression procedure is employed to estimate the model. The stepwise estimation procedure is used because in the end we want to obtain a manageable number of variables that can be relatively easily collected in practice and at the same time meaningfully used to predict household consumption level and poverty status.

3.3 The Augmented model

The usual approach concerning poverty measurements has historically been to model poverty Directly.The consumption model can be described as the basic model.Futhermore the model of consumption c_j, the determinants of per capita consumption at the household level in the simplest form of a model is as follows

$$logc_j = \boldsymbol{\beta}_\mathbf{j}\mathbf{x}_\mathbf{j} + e_j \tag{1}$$

where x_j is a set of household characteristics and e_j is a random error term.. It has the feature that the marginal effects of the determinants of consumption are constant across households. It is however arguable that there is heterogenity across households and the marginal effects themselves depend on household characteristics. This concern leads us to consider the augmented model that allows for a range of interaction effects and individual specific marginal effects ($\boldsymbol{\beta}_\mathbf{j}$);

$$logc_j = \boldsymbol{\beta}_\mathbf{j}\mathbf{x}_\mathbf{j} + e_j$$

where $\boldsymbol{\beta}_\mathbf{j} = \boldsymbol{\beta}' + \mathbf{x}_\mathbf{j} + \mathbf{e}_\mathbf{j}$ and hence

$$logc_j = \boldsymbol{\beta}'\mathbf{x}_\mathbf{j} + x_j\phi x_j + e_j^* \tag{2}$$

This delivers a model with heteroscedastic errors, $e_j^* = e_j + \varepsilon_j$, which is easily allowed for estimating the variance matrix of the model parameters. The model has a generalized quadratic form which is a numerically equivalent second order approximation to any arbitrary twice differentiable function (Fahrmeir and Kaufmann, 1985). [25].

3.4 Specification of the Poverty logistic Model

Choosing an appropriate model and analytical technique depends on the type of variable under investigation. Regression deal with cases where the dependent variable of interest is a continuous variable which we assume, perhaps after an appropriate transformation, to be normally distributed. But in many applications, the dependent variable of interest is not on a continuous scale; it may have only two possible outcomes and therefore can be represented by an indicator variable taking on values 0 and 1.

In this study, the dependent variable Y was defined to have two possible outcomes:

1. The household is poor (1)
2. The household is not poor (0)

These two outcomes are coded 1 and 0 respectively. This shows that the dependent variable is dichotomous and it can be represented by a variable taking the value 1 with probability π and the value 0 with probability 1−π. Such a variable is a point binomial variable, that is, a binomial variable with n =1 trial, and the model often used to express the probability π as a function of potential independent variables under investigation is the logistic regression model. Therefore, to sort out which explanatory variables are most closely related to the dependent variable, nine factors are considered. This method involves a linear combination of the explanatory or independent variables. Thus, the study is modeled within the framework of above mentioned theories and the model used by this study to determine factors affecting poverty status is given equation (3).

3.5 Logistic Regression Analysis

The function has been discussed by many researchers like [26]. It is given by;

$$f(g) = \frac{exp(g)}{1+exp(g)} = \frac{1}{1+exp(g)} \tag{3}$$

when modeling a Bernoulli random variable with multivariate, one directly models the probabilities of group membership, as follows;

$$P(Y=1 \mid X=x) = \frac{1}{1+exp\left(-\left(\beta_0 + \sum_{j=1}^{d} x_j \beta_j\right)\right)} \tag{4}$$

where g in Equation 3 is given by

$$g(X;\beta) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_d X_d \tag{5}$$

To illustrate, the applicability of the logistic function, the bold curve in the figure 2 shows that the logistic function puts more weight on the tails than the normal distribution.

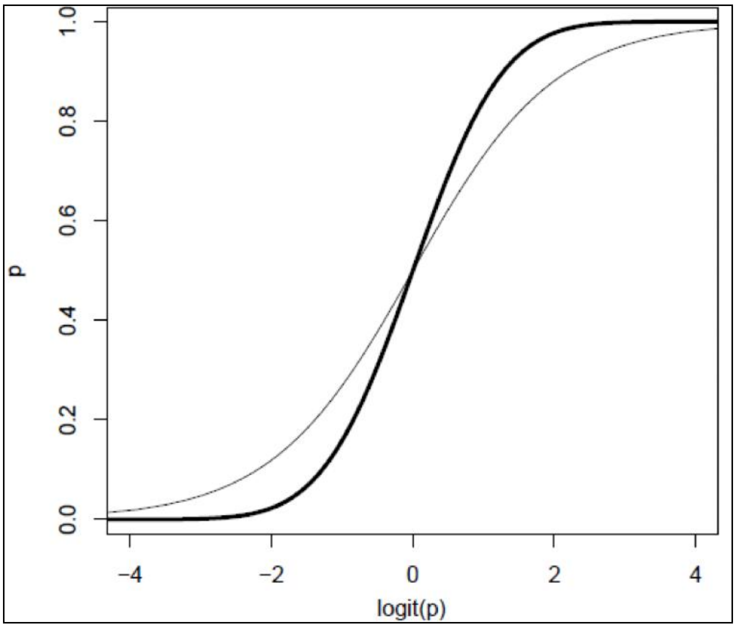

Figure 2: Standardized Normal and Logistic CDF's

Author (2014)

The logistic model is bounded between zero and one, this property estimates the possibility of getting estimated or predicted probabilities outside this range which would not make sense. Also with a proper transformation, one can get a linear model from the logistic function. [26] uses the logit function of the Bernoulli distributed response variable. Transforming Equation 4 as in [26] we have ;

$$Logit[P(Y=1 \mid X=x)] = log_e \frac{P(Y=1 \mid X=x)}{1-P(Y=1 \mid X=x)}$$

$$= log_e \left\{ \frac{1 + exp\left(\beta_0 + \sum_{j=1}^{d} \beta_j X_j \right)}{1 + exp\left(-\left(\beta_0 + \sum_{j=1}^{d} \beta_j X_j \right)\right)} \right\}$$

$$= log_e \left(exp\left(\beta_0 + \sum_{j=1}^{d} \beta_j X_j \right)\right)$$

$$= \beta_0 + \sum_{j=1}^{d} \beta_j X_j \qquad (6)$$

the function in Equation 6 is a generalized linear model (GLM) with d independent variables.

The motivation to the use of logistic model is that it follows the properties of the GLM. Lets define the hypothetical population proportion of cells for which $Y = 1$ as $\pi = P(Y = 1 \mid X = x)$. Then the theoretical proportion of cells for which $Y = 0$ is $1 - \pi = P(Y = 0 \mid X = x)$. We estimate π by the sample proportions of cells for which $Y = 1$. In the GLM context, it is assumed that there exists a set of predictor variables, $X_1, X_2, \cdots, X_d$, that are related to Y and therefore provides additional information for estimating Y. For mathematical reasons of additivity and multiplicity, logistic model is based on linear model for the log odds in favour of $Y = 1$.

$$log_e \frac{\pi_i}{1 - \pi_i} = \alpha + \sum_{j=1}^{d} \beta_j X_j \qquad (7)$$

thus

$$\pi_i = \sum_{j=0}^{d} \beta_j X_j \qquad (8)$$

where $\beta \in \Re^d$ of unknown parameters.
The logistic regression (logit link),

$$g(\pi_i) = log_e \frac{\pi_i}{1 - \pi_i} = logit(\pi_i)$$

and

$$g^{-1}(g(\pi_i)) = \pi_i$$

thus the inverse of the logit function in terms of $(X; \beta)$ is given by;

$$g^{-1}(X; \beta) = \pi_i = \frac{1}{1 + exp\left(-\left(\beta_0 + \sum_{j=1}^{d} x_j \beta_j \right)\right)}$$

This model can be rewritten as

$$logit(\pi_i) = \sum_{j=0}^{d} \beta_j X_j \qquad (9)$$

IV. Results and Discussion

4.1 Empirical studies of the Stepwise Regression Model and the Augmented Regression Model

For several of the explanatory variables, there are observations with missing data and have constructed dummy variable that take a value of one if the household is missing data for a particular variable(while the value of that variable itself is set as zero). In this way, we reduce the potential of sample selection bias, and we do not miss out on useful information from household with some valid data for most variables.

Per capita consumption is used as the basic measure of individual welfare. The use of per capita consumption imposes the assumptions that there are no economies of household size in consumption and that household composition does not matter, and therefore, the estimated parameters must be intepreted with caution.

There may also be some concern of potential bias in parameter estimates due to endogeneity of omitted variables. If these factors are significant determinants of welfare, the error term will not converge to zero in probability limit and the parameter estimated for the individual explanatory variables will be inconsistent. To control this, interactions term effects are included in the model.

While the augmented equation 2 offers a fairly general approach to modeling welfare, this generality comes with the potential cost of overparameterizing the model with the full set of interaction terms, there are an explosion of parameters. Beginning with a k-parameters in the basic model, there are $\frac{2k+k(k-1)}{2}$ parameters in the augmented equation 2 .

A model with numerous parameters is likely to suffer from multicollinearity. In the view of these difficulties; we use the stepwise regression as our basic model so as to limit them to only those significant in the model. see Micheal. H.K et al.(2005),[39]

Table 2: Stepwise and augmented modeling of the log per capita consumption

Variables	Description	Stepwise model		Augmented model	
		Coefficient	t-ratio	Coefficient	t-ratio
X1	Hhsize	0.4079(.)	1.963	2.297(***)	5.466
X2	Hh size ^{2}	-0.028(*)	-2.062	-0.2410(**)	-3.356
X3	Gender Hh (head)	0.4988(.)	1.853		
X5	Land size (acre)	0.5824(.)	1.983	0.5335(*)	2.203
X6	Hh (head)age	0,1588(***)	3.575		
X7	Hh (head) age ^{2}	-0.0016(**)	-3.575		
X8	Hh Aveage in school	0.0857(.)	1.868	0.2568(**)	2.886
X9	Production(kg) per year	0.0005(***)	1.890	0.0029(*)	2.676
X1:X8	Hh size* Hh Aveage in school			-0.0277(*)	-2.139

(Significance codes: *** 0.001 , **0.01 , *0.05 . 0.01)

(Hh-Household: A domestic unit consisting of members of a family who live together)

Table 2 represents both the stepwise regression model and the augmented model. The null hypothesis, that interactions in the augmented model are jointly equal to zero is convincingly rejected. Thus, there is no support for the standards are uniform across households.

The household size has significant negative (though nonlinear) effects on welfare. This inverse relation between household size and the log per capita consumption is a common finding in the literature (Lanjouw and Ravallion, 1995; Lipton, 2001), [27,28]. The measure of per capita consumption as used in the study is the total food consumption ,non-food and othe expenses of the household. Each of these components of consumption is well documented in more details in the basic report of well-being in Kenya 2005/06 .thus consumption is critically dependent on the underlying assumption regarding economies of household size and equivalent scales.

Education variable emerge as a strong determinant of welfare. In both models the average years of schooling specified on its own have significant positive effects on per capita consumption. However, once the models have been augmented with interactions, several interaction terms in schooling are found to be significant. For example, the marginal return to school is found to be increasing with household size as well as decreasing with the number of the years in school.

We find a strong positive significance effects on the average number of years in education for the family. The models indicate strong positive effects on household if the family is educated. Oduro et al. (2004) ,[29] argue that education and skill acquisition are critical factors for explaining the pattern of rural poverty. Education contributes to the process of moulding attitudinal skills and developing technical skills, and also facilitates the adoption and modification of technology [29].

The study finds that family that owned land (for production) has a significant positive effect on per capita consumption of the household, The age of the household head shows that the expected life cycle in the stepwise model increases poverty status by 15%, also the quadratic term of the age which is nonlinear shows a decline in the life cycle phenomenon of high earning capacity with greater experience and smoothing of consumption over life cycle. There have been similar finding by other authors though using a different techniques, (Datt and Jolliffe, 2005; Mwabu et al., 2000; Oyugi, 2000), [7,30,31].

Table 3: Suitability of the models as indicator of poverty

Models	R^2	Standard error
Stepwise	0.9917	0.815
Augmented	0.9946	0.6895

4.2 Empirical studies of the Logistic Model

This method predicts poverty directly because of the nature of the dependent variable. There are two things that need to be reiterated. First, the dependent variable takes values the values of 1 when the respondent is poor and 0 otherwise.This means in interpreting the estimation result it is important to remember that a positive coefficient means that the variable is correlated positively with the poor. Second, predicted value of the dependent variable is the probability of the observation to be poor.

A logit model has been estimated to elicit the factors influencing welfare status of households. The model uses current welfare status of household as the dichotomous dependent variable. poverty variable is defined on the basis of the variable determinant of poverty indicated below.

The variables in this case are:

Y_i	Poverty of household i ($1=$ Poor, and $0=$ Non-Poor)
X_1	Household size
X_2	Square of household size
X_3	Gender of household head ($1=$ male, and $0=$ female)
X_4	land size(acres)
X_5	Education of HH head ($1=$Primary level and above, $0=$No Education)
X_6	Age of Hh (head)
X_7	Square of Age of Hh (head)
X_8	Per capita aggregate production (No. of Kgs)

The logistic model was fitted to the data to test the relationship between the likelihood of a household being poor or non-poor. The logistic regression analysis was carried out by stepwise method , and the result showed that

The optimal model

$$\begin{aligned} Z_R &= -1.4721\ X_1 + 0.1398\ X_2 + 1.6905\ X_3 + 0.0358\ X_4 \\ &+ 0.0781\ X_5 + 0.3796\ X_6 - 0.0059\ X_7 - 0.3659\ X_8 \end{aligned} \tag{10}$$

According to the model, equation 10, the log of the odds of a household being poor was negatively related to size of the household p=0.01), which according to literature, Paddy (2003) [31] noted that household size was negatively correlated to poverty and Deaton and Paxson (1995) [32] found that food requirement increased in relation to the number of persons in household. The non-linear component of the household size is positively correlated to poverty. This is a common finding in the literature, see [27] and [28].

The log of odds of the gender of the head of the household was positively related to the poverty $(p = 0.05)$.

The age of the household head shows the expected life expectancy. In our model, household living standard increases with the age of the household head upto the optimal age of around 60 years but decreases with the quadratic term which is significant $p = 0.05$. This is consistent with higher earning with greater experience.

There is a strong intergenerational effect on education. Parental education has a strong positive correlation on household welfare.

Food production was expected to be increased extensively through expansion of areas under utilization. The model indicates land size increased food security with 0.4290 even though $(p > 0.05)$. The model, the log odds of land size in positively related to poverty $(p = 0.05)$. In other words, the larger the size of land the increase to production. The production (kg) of the household, the log of the odds indicates that a unit increase of food production improved the food poverty status of the household by 1.4%, with $(p > 0.05)$.

Table 4: Predictors

Predictors	β	$SE(\beta)$	z	p **-value**	e^{β} **(Odds ratios)**
Size of Hh (numbers)	-1.4721	0.0907	-16.230	0.1090	0.2294
Square of household size	0.1398	0.0810	1.7235	0.0844	1.1500
Gender of Hh head (1-male, 0-female)	1.6905	0.8790	1.9232	0.0545	1.0560
Land size(acres)	0.0358	0.2449	0.1461	0.8836	2.4196
Education of Hh head ($1 =$ Primary level and above, $0 =$ No Education)	0.0781	0.1054	0.7419	0.4587	1.5820
Age of of Hh head	0.3796	0.1768	2.1670	0.0318 *	1.0000
Square of Age of Hh head	-0.0059	0.0026	-2.2192	0.0257 *	1.0260
Per capita aggregate production (No. of Kgs)	0.3659	0.1671	2.1897	0.0139^{*}	1.0140

(Dispersion parameter for binomial family taken to be 1)

4.3 Evaluation of the logistic regression model

The overall model evaluation is said to provide a better fit to the data if it demonstrates an improvement over the intercept only model (also called the null model). An intercept only model serves as a good baseline because it contains no predictors. According to this model, all observations would be predicted to belong in the largest outcome category. An improvement over this baseline is examined by using three inferential statistical tests.

Table 5: Statistical inference table Statistical test

Statistics Test	χ^2	df	p
Likelihood ratio test	9.0353	5	0.0854
Hosmer-Lemeshow	9.6702	5	0.7418
Wald test	4.0456	5	0.5443

The statistical significance of individual regression coefficient i.e. (β 's) is tested using the Wald chi-square statistic. According to table 5, the variables are significant predictors of poverty $(p < 0.05)$.

Goodness-of-fit statistics assess the fit of a model against actual values. The inferential goodness-of-fit test is the Hosmer-Lemeshow (H-L) test that yields a $\chi^2_{(5)}$ of 9.6702 and was insignificant $(p < 0.05)$. Suggesting that the model fits the data well. In other word's, the null hypothesis model of a good model fit to data was tenable. The likelihood ratio test yields a $\chi^2_{(5)}$ of 9.0353 and was significant at $p > 0.05$ which also give a good fit for the model and thus the null hypothesis was also tenable for the model.

Table 6: 95% confidence interval for one unit change in X_t

Size of Hh (Number)	$-3.7940, 0.0835$
Square of household size	$0.0002, 0.3473$
Gender of Hh head (1-Male, 0-Female)	$0.1099, 3.8349$
Land size	$-0.4561, 0.3260$
Education of Hh head ($1 =$ Primary level and above, $0 =$ No Education	$-1.0770, 2.5661$
Age of Hh head	$0.0930, 0.8789$
Square of Age of Hh head	$-0.0120, 0.0014$
Per capita aggregate production (kg)	$-0.7457, -0.1033$

The full model is:

$$\begin{aligned} Z_F &= -2.5237\ X_1 + 0.2237\ X_2 + 2.4270\ X_3 + 0.0358\ X_4 \\ &+ 0.7810\ X_5 + 0.6391\ X_6 - 0.00932\ X_7 - 0.3834\ X_8 - 2.076\ X_{11} - 0.0024\ X_{13} \end{aligned} \quad (11)$$

We wish to test

$$H_0 : \beta_0 = \beta_1 = \beta_2 = \cdots \beta_{10}$$

$$H_A : \beta_j \neq 0$$

The reduced model is:

$$\begin{aligned} Z_R &= -1.4721\ X_1 + 0.1398\ X_2 + 1.6905\ X_3 + 0.0358\ X_4 \\ &+ 0.0781\ X_5 + 0.3796\ X_6 - 0.0059\ X_7 - 0.3659\ X_8 \end{aligned}$$

Table 7: Deviance analysis of the model

Model	Null Deviance	df	Residual Deviance	df
Full model	66.542	48	40.373	35
Reduced model	44.317	32	24.405	24

Therefore, we do not reject the hypothesis, and conclude that the reduced model is a better model than the full model.

4.4 Comparison of the two models using the confusion matrix

The confusion matrix is commonly used to compare two models on how good the predicted respondents. In our study the following matrix were obtained:

Table 8: Logistic model

Indicator observed		
	1	**0**
1	35	0
0	0	23

Table 9: Augmented model

Indicator observed		
	1	**0**
1	34	0
0	0	23

The confusion matrix informs us that the logistic model is better for predicting poverty than the augmented model since it has a high prediction of accurate respondents than the augmented.

4.5 Housing conditions

4.5.1 Roofing Material as measure of poverty

Majority of the respondent represented by 78% stay in corrugated iron sheet houses, followed by with glass thatched houses at 26%, there are also about 1% houses roofed with tiles, another 2.5% with asbestos and the other with about 3% roofed by other materials, this factor may not give a good indicator of poverty but if looked from the perspective of the whole house building material we will be able to see that this indicator can be able to give some indication of poverty.

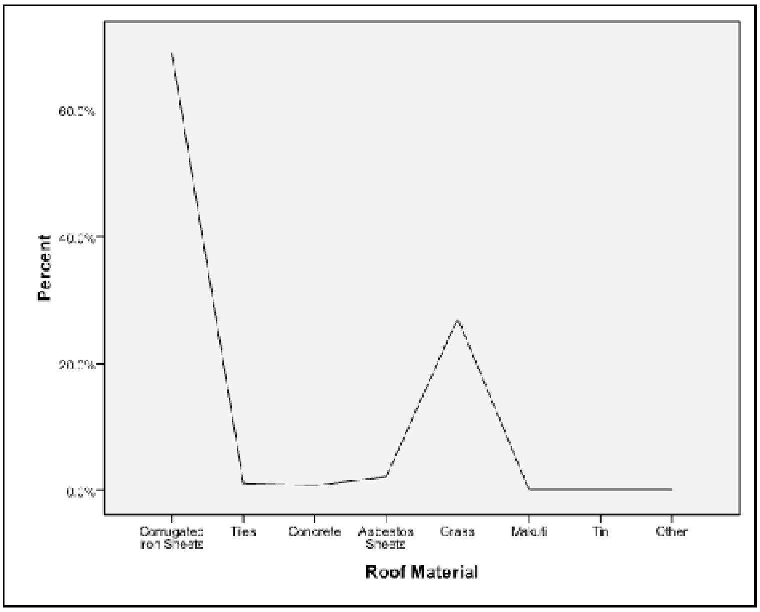

Figure 3: Roofing Materials

4.5.2 Wall material as a measure of poverty

The Majority of houses are walled using mud and wood which represents 62%, 19% are made of bricks, 17% are walled with mud and cement and the others about 3% are walled with other materials like timber and stone which indicates that even combined with roofing materials this area poverty is very high.

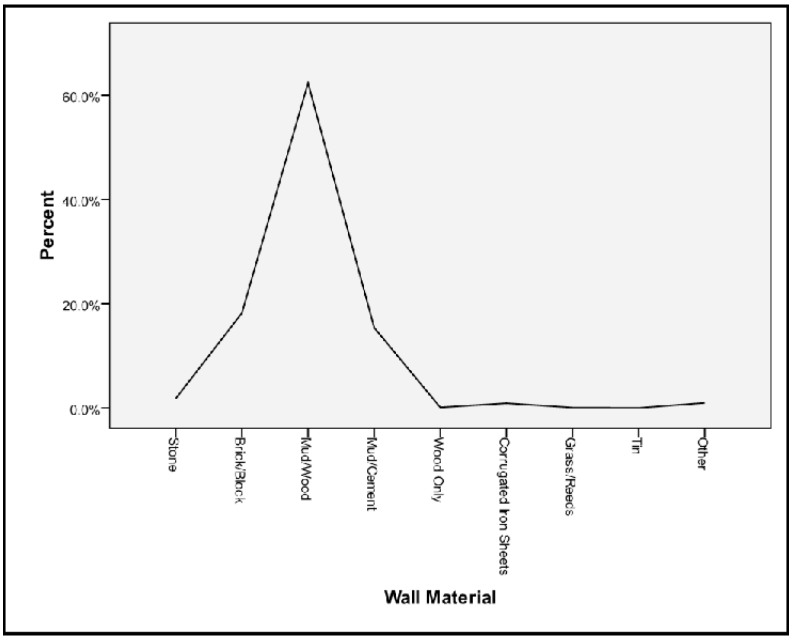

Figure 4: Wall Materials

4.5.3 Main water sources as a measure of poverty

According to [8] about 1.1 billion people lack access to improved water sources, which represents 17% of the global population. In order to achieve the millenium goals, many efforts needs to be done in the areas to ensure the people have clean and safe water. The area majority about 80% only get water from rivers, lake and streams which many times are not clean. [33] also argues that limited access to basic services such as to running water, sanitation on site, grid electricity and health care services is an impediment to escaping from poverty.

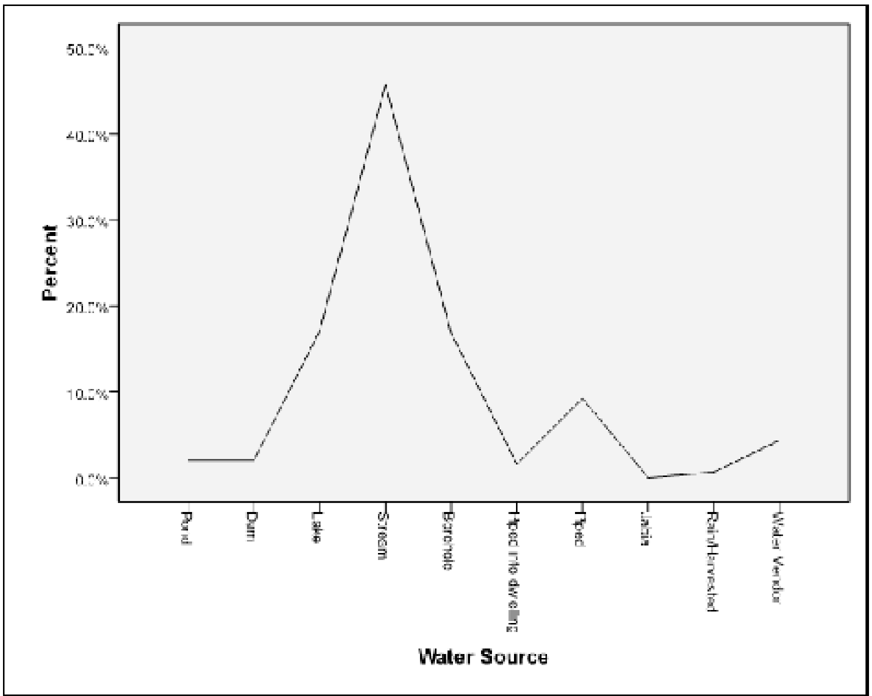

Figure 5: Water Sources

4.6 Information regarding livestock

4.6.1 Poverty against Indigenous cattle

Table 8: Indegenous cattle Chi-square Test

Tests	Value	d.f	Asy. Significance
Pearson chi-square	155.835 [a]	13	0.000
Likelihood ratio test	152.524	13	0.000
Linear by Linear association	73.569	1	0.000
N of valid cases	4414		

[a] 0 cells (0%) have expected count less than 5. The minimum expected count is 7.51

In the table 7, we can see that $chi-squared\ test\ (13) = 155.835$ at $p < 0.05$. Since the p-value is less than 0.05, we reject the null hypothesis and say that there is statistically significant association between poverty and the rearing of the indigenous cattle in the region. The sample size requirement for chi-squared test of independence is satisfied since zero cells (0 %) has expected count less than 5.

4.5.2 Poverty against Goat

Table 9: Goat Chi-square Test

Tests	Value	d.f	Asy. significance
Pearson chi-square	85.213 [a]	11	0.000
Likelihood ratio test	82.091	11	0.000
Linear by Linear association	45.524	1	0.000
N of valid cases	4414		

[a] 0 cells (0%) have expected count less than 5. The minimum expected count is 5.23

The table 8, shows the relationship between poverty and goat rearing is also statistically significant as we can see from the $chi-squared\ test(11) = 85.213$ at $p < 0.05$.The sample size requirement for chi-squared test of independence is satisfied since zero cells (0 %) has expected count less than 5.

4.5.3 Poverty against Sheep

Table 10: Sheep Chi-square Test

Tests	Value	d.f	Asy. significance
Pearson chi-square	30.444 [a]	8	0.000
Likelihood ratio test	29.185	8	0.000
Linear by Linear association	16.543	1	0.000
N of valid cases	4414		

[a] 0 cells (0%) have expected count less than 5. The minimum expected count is 9.80

Table 9 indicates also that in the region there exists a relationship between poverty and sheep rearing which is statistically significant with $chi-squared\ test\ (8) = 30.444$ at $p < 0.05$.The sample size requirement for chi-squared test of independence is satisfied since zero cells (0 %) has expected count less than 5.

Number of total livestock units owned reduce household poverty rates, implying that assets are important determinants of poverty. This finding is consistent with earlier findings for Kenya [20, 21 and 30].

V. Conclusion

The main objective of the study even with difficulties of obtaining expenditure and income data household precise data and to finding variables that predict poverty in rural areas of Kenya is achieved. In the study we explore the two methods, augmented regression model and the logistic regression model, on predicting poverty .The logistic model was better since it was able to predict correctly all respondents, while the augmented model had a prediction rate of about 2% of not predicting correctly the respondents in the consumption model. However, since our aim is to predict the poor for policy mitigation we focus on the method that provides us with the most accurate prediction. In predicting the poor the logistic model is the best of the consumption models.

Further, we also notice that the variables with the strongest either positive or negative are Land, education, size of the household, age of household head and gender. Furthermore, house characteristics, access to facility and assets play significant role. Thus, if we want to roughly assess whether a household is more likely to be poor or not in the region, it would be better to gather information on assets ownership, education level and consumption patterns as they are the best indicators that should be used to tell the status of poverty in a household. Considering the current population growth rate of about 2.5 percent per annum, there is need for a general overview of the policies to boost economic growth and measures to ensure reduction of poverty to the majority of Kenyans. This should be combined with promotion of family planning to ensure that economic gains and reduced burden on households, as a result of free or subsidized services (e.g. in education and health), do not translate to higher population growth. There is also need for targeted investments in infrastructure such as roads, rural electrification, safety net programmes and provision of water, especially in the marginal areas. The policies on poverty levels in the lake region under the PRSP's three pillar strategy of raising the income opportunities for the poor should focus mostly on agriculture, since the macroeconomic environment is important in determining the productivity which is key to poverty reduction.

References

[1]. World Bank (WB), "World development indicators 2010," Technical report License: CC BY 3.0 IGO, World Bank, Washington, DC, 2010. Available at: https://openknowledge.worldbank.org/handle/10986/4373.

[2]. S. Pudney, "On some statistical methods for modelling the incidence of poverty," Oxford Bulletin of Economics and Statistics, vol. 61, no. 3, pp. 385–408, 1999.

[3]. M. Ravallion and B. Bidani, "How robust is a poverty profile?," World Bank Economic Review, vol. 8, no. 1, pp. 75–102., 1994.

[4]. N. Kakwani, "Poverty and economic growth with an application to cote d'lvoire," LSMS Working Paper 63, World Bank, Washington, D.C, 1990.

[5]. M. Lipton and M. Ravallion, Handbook of Development Economic, vol. 3 of B. Amsterdam: Elsevier, 1995.

[6]. A. Sen, Development as Freedom. Oxford: Oxford University Press, 1999.

[7]. G. Datt and D. Jolliffe, "Poverty in egypt: Modeling and policy simulations," Economic Development and Cultural Change, vol. 53, pp. 327–346, January 2005 2005.

[8]. WHO, "Energy and protein requirements," Tech. Rep. 724, World Health Organization, Geneva, 2000.

[9]. M. Ravallion, ""poverty lines in theory and practice",living standards measurements surveys," tech. rep., World Bank, 1998.

[10]. A. M. P. Prescott, Nicholas, "a poverty profile of cambodia," Discussion Paper 373, World Bank, Washington, DC, 1997.

[11]. KNBS, "Basic report on well being in kenya based on kenya integrated household budgets survey, 2005/2006," tech. rep., Kenya National Bureau of Statistics, 2007.
[12]. KNBS, "Geographic dimension of well being in kenya," tech. rep., Kenya National Bureau of Statistics, 2007.
[13]. Schreiner, Mark. "A Simple Poverty Scorecard for Tanzania", http://www.microfinance.com/English/Papers/Scoring_Poverty_Tanzania EN_2013.pdf, retrieved 31 January 2015..
[14]. P. Kotler, N. Roberto, and T. Leisner, "Alleviating poverty: A macro/micro marketing perspective," Journal of Macromarketing, vol. 26, no. 3, pp. 233–238, 2006.
[15]. C. Laderchi, R. Saith, and F. Stewart, "Does it matter that we do not agree on the definition of poverty: A comparison of four approaches.," Oxford Development Studies, vol. 31, no. 3, pp. 233–274, 2003.
[16]. M. Rank, One Nation Underprivileged: Why American Poverty Affects Us All. New York, NY: Oxford Press., 2004.
[17]. B. Arne, Regional inequality in Kenya. Heinemann Eduacational Books Ltd., 1981.
[18]. A. Hazlewood, 'Income Distribution and Poverty;An Unfortunate View' in Tony Killick. Heinemann Eduacational Books Ltd, 1981.
[19]. W. House and T. Killick, 'Inequality and Poverty in Rural Economy and the Influence of some Aspects of Policy'. Heinemann Eduacational Books Ltd, 1981.
[20]. A. Geda, N. de Jong, G. Mwabu, and M. S. Kimenyi, "Determinants of poverty inkenya: Household-level analysis," Discussion Paper Series 9, The Kenya Institute for Public Policy Research and Analysis (KIPPRA), Nairobi, 2001.
[21]. L. N. Oyugi, "The determinants of poverty in kenya," Master's thesis, Department of Economics , University of Nairobi, 2000. MA Thesis, Department of Economics ,University of Nairobi.
[22]. J. Greer and E. Thorbecke, "A methodology for measuring food poverty applied in kenya," Journal of Developement Economics, vol. Vol. 24, pp. pp. 59–74, 1986.
[23]. J. Greer and E. Thorbecke, "Food poverty profile applied in kenya smallholders," Economic Development and Cultural Change, vol. 35, no. 1, pp. 115–141, 1986.
[24]. P. Mwita, V. Masanja, C. Muyanja, and R. Odhiambo, "Mathematical techniques for food crops balance sheet and food security indicators in lake victoria watershed," tech. rep., Inter-University Council of East Africa, Arusha, 2007.
[25]. I. Fahrmeir and H. Kaufmann, "Consistency and asymptotic normality of maximum likelihood estimator in generalized linear models," Annals of Statistics, vol. 13, pp. 342–368, 1985.
[26]. J. Fan, M. Farmen, and J. Gijbels, "Local maximum likelihood estimator and inference," Stat. Soc. Ser. B Stat. Methodol., vol. 60, no. 3, pp. 591–608, 1998.
[27]. P. Lanjouw and M. Ravallion, "Poverty and household size," The Economic Journal, vol. 105, no. 433, pp. 1415–1434, 1995.
[28]. M. Lipton, "Challenges to meet: food and nutrition security in the new millennium," Nutrition Society, vol. 60, no. 2, pp. 203–214, 2001.
[29]. A. D. Oduro, I. Osei-Akoto, and I. Acquaye, "Poverty in a globalizing economy: The role of rural institutions," tech. rep., FASID, Japan, 2004.
[30]. G. Mwabu, W. Masai, R. Gesami, J. Kirimi, F. Munene, M. Chemengich, and J. Mariara, "Poverty in kenya; profiles and determinants," tech. rep., University of Nairobi and Ministry of Finance & Planning, Mimeo, Nairobi, 2000.
[31]. F. Paddy, "Gender differentials in land ownership and their impact on household food security: A case study of masaka," Master's thesis, Uni–Hohenheim, Uganda, 2003. http:www.troz.uni–hohenheim.de/research/thesis/MSCAES/paddy.pdf.
[32]. A. Deaton and C. Paxson, "Measuring poverty among the elderly," tech. rep., National Bureau of Economic Research, 1995.
[33]. S. Van der Berg, "Poverty and the role of rural institutions in a globalizing south african economy," tech. rep., FASID, Japan, 2004.
[34]. Datt G & Jolliffe D , Poverty in Egypt: Modeling and Policy Simulations, Economic Development and Cultural Change, Vol. 53, pp. 327–346, January 2005
[35]. Ravallion, M., Chen, S., and Sangraula, P. (2007). New evidence on the urbanization of global poverty. Population and Development Review, 33(4):667–701.
[36]. Collier, P. (2007). The bottom billion: Why the poorest countries are failing and what can be done about it. Oxford University Press.
[37]. Jamison, D. T., Breman, J. G., Measham, A. R., Alleyne, G., Claeson, M., Evans, D. B., Jha, P., Mills, A., and Musgrove, P. (2006). Disease control priorities in developing countries. World Bank Publications.
[38]. Arrow, K., Dasgupta, P., Goulder, L., Daily, G., Ehrlich, P., Heal, G., Levin, S., Mäler, K.-G., Schneider, S., Starrett, D., et al. (2004). Are we consuming too much? Journal of Economic Perspectives, pages 147–172.
[39]. Micheal, K. H. and Li, W. (2005). Applied Linear Statistical Models. McGraw-Hill, Irwin, 5th edition.