DEPARTMENT OF STATISTICS

UNIVERSITY OF GHANA

PREDICTING POVERTY INCIDENCE IN STATISTICALLY UNDERDEVELOPED COUNTRIES

BY

ERNEST ENYAN
ID: 10042177

MPHIL STATISTICS

THIS THESIS IS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES, UNIVERSITY OF GHANA, IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF PHILOSOPHY DEGREE IN STATISTICS

JUNE, 2015

DECLARATION

Candidate's Declaration:

I, hereby, certify that the work presented in this thesis is my own and that it has not partially or wholly been presented for another degree in this University or elsewhere. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

SIGNATURE…………………………..   DATE……………………………

ERNEST ENYAN

(10042177)

Supervisors' Declaration

We, hereby, certify that this thesis was prepared from the candidate's own work and supervised in accordance with guidelines on supervision of thesis specified by the University of Ghana.

SIGNATURE……………………………   DATE………………………..

DR. ISAAC BAIDOO

(Principal Supervisor)

SIGNATURE…………………………..   DATE…………………………..

DR. ANANI LOTSI

(Co – Supervisor)

ABSTRACT

It is generally acknowledged that availability of data remains critical to effective monitoring of poverty indicators and evaluation of policies and programmes towards alleviating poverty. The paucity of data has reduced the frequency of measuring poverty indicators at the national and regional level. District level poverty indicators are yet to be estimated since sampling designed data collected can only provide national and regional level indicators. This thesis applied multiple imputation technique to predict poverty incidence from a non-expenditure household data (MICS4) based on a regression model developed from a recent household expenditure data (GLSS6). The time interval between the two data sets is one and a half years, and few variables in GLSS6 data that significantly influence poverty were not found in the MICS4 data. However, the poverty incidence predicted for national, urban and rural levels at 95 percent probability were very close to that estimated by the Ghana Statistical Service from the GLSS6 data. The study recommends application of this technique to future non-expenditure survey to improve the frequency of poverty indicators to inform policy and programmes.

DEDICATION

This work is dedicated to my lovely wife Anita Antoh, my precious children Daniel, Julia, Nkunim and Adepa for their prayers and support throughout this course. I will forever be grateful to you all.

ACKNOWLEGDEMENT

I acknowledge with sincere thanks to my lecturers and supervisors Dr. Isaac Baidoo and Dr. Anani Lotse for their invaluable encouragements, guidance and suggestions without which this work would not have been completed. I also thank Dr. Samuel Idi and Dr. F. O. Mettle for their countless suggestions and directions.

I am particularly grateful to the government statistician, Dr. Philomena Nyarko, for her encouragement and the study leave with pay granted me to undertake this course. I acknowledge with gratitude the financial support from the World Bank through Ghana Statistical Service to pay for my second year fees and also to purchase certain books for the course. I give special thanks to my unit head, Mrs. Jacqueline Anum who gave me the opportunity to meet with Dr. Dhiraj Sharma and Dr. Nobuo Yoshida from the World Bank for technical support. In addition, I am sincerely grateful to Dr. Dhiraj Sharma who relentlessly inspired and supported me with technical advice.

I thank my mother for her prayers, my siblings for their support and all colleagues as well as course mates for inspiring me in one way or the other to undertake this course successfully. Finally, I would like to acknowledge the extraordinary support of my family for the invaluable assistance that has contributed in diverse ways to the successful completion of this course.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION/BACKGROUND

1.0 INTRODUCTION

Ghana's Poverty profile, 2012/13, estimated that 24.2 percent out of the total population of about 24.6 million (GSS, 2013) were living in absolute poverty. Consequently, poverty reduction interventions will continue to be carried out in Ghana, even into the future until poverty is completely alleviated or reduced to the barest minimum. An important part of this effort to reduce poverty is monitoring and evaluation of projects and programmes towards the poor, which is required to be regular and systematic.

An effective poverty monitoring will detect signs of deterioration or improvement in people's welfare, so that the relevant authorities can take appropriate and timely action to prevent the situation from worsening or sustain and advance the progress made. Such monitoring requires regular and frequent estimations of various welfare indicators. Hence, a regular and frequent effort to collect household level data is a crucial part of an effective poverty monitoring system.

The most widely used data for measuring welfare or poverty is household consumption expenditure. The advantage of using household expenditure data is that, it is already expressed in monetary units, which relatively makes calculating a poverty line and poverty headcount rate straightforward.

Collecting household consumption expenditure data requires a long period of time and much effort. It requires respondents' commitment and willingness to keep record of their expenditure for a stipulated period, about the past 12 months, and the enumerators' sufficient trust that the respondents will correctly record their actual expenditure for the period required. Furthermore, if the questions asked require the memory of respondents in recalling their food and non-food expenditure over the past 12 months, then questions accuracy and reliability become important to consider. In the face of these difficulties, a number of studies such as the Core Welfare Indicators Questionnaire (CWIQ), especially in developing countries, have been conducted to try and address this empirical problem by creating a proxy measure of poverty. The proxy is calculated using widely recognized methodologies and household characteristics such as assets ownership, household size and education which are comparatively easier to collect, and have been proven to significantly influence poverty. The main purpose of using the proxy is to obtain an estimate that places households at the same level they would be, had they been graded using per capita consumption expenditure. Additionally, the study is to ascertain which of these methodologies give the best estimates that would place the households at the same level they would have been if graded by per capita consumption expenditure.

1.1 BACKGROUND

The well-being of individuals, households and communities as well as countries has called for many interventions and programmes, including the millennium development goals of which the goal one requires member countries of the United Nations to half poverty and hunger by the end of 2015. According to the World Bank (2000), poverty is the most important defining characteristic of underdevelopment, and national poverty rate is the core measure of living standards drawing attention exclusively towards the poor. National estimates of poverty rate (i.e. poverty line) are based on population-weighted subgroup estimates derived from household surveys.

An individual or a household is considered poor if the consumption or income levels the individual or the household falls below some minimum level necessary to meet basic needs. This minimum level is usually called the "poverty line". However, what is necessary to satisfy basic needs varies across time and societies. Poverty lines vary in time and place, and each country uses lines which are appropriate to its level of development, societal norms and values.

Gönner C., Haug M., Cahyat A., Wollenberg E., De Jong W., Limberg G., Cronkleton P., Moeliono M. and Becker M., (2010), stated that poverty concepts have profoundly changed from the mere consideration of income or consumption, to definitions that include multiple dimensions of deprivation and well-being. The authors emphasized that currently, leading development organisations apply poverty definitions that comprise aspects like self-determined lifestyles, choice, assets, capabilities, social inclusion, inequality, human rights, entitlement, vulnerability, empowerment and subjective wellbeing. In the authors view, the new poverty concepts have found their way into the UN Human Development Report (UNDP, 2005), the World Bank’s World Development Report (2000, 2001 and 2002) and into other more qualitative poverty studies published by the World Bank. The authors stated that these new concepts, which are more sophisticated, have been difficult to quantify, therefore, international agencies such as the World Bank and UNDP as well as national governments, still favour money-metric poverty lines like the famous Purchasing Power Parity (PPP) index of $1 or $1.25 per day or non-fulfillment of basic needs.

Ghana, since independence, has rolled out many programmes and interventions to combat the multi-faceted phenomenon of poverty. In the last three decades, a series of six (6) Living Standard Surveys (GLSS I, II, III, IV, V and VI) have been conducted by the Ghana Statistical Service where each round of the survey covered a nationally representative sample of households, spread over the entire country, with data collection period of 12 months. These surveys also covered a wide range of household characteristics and behaviour, and measured the welfare of residents in the country to inform government policies and programmes.

The Ghana Living Standards Surveys (GSS, 1991-2006) acknowledged that poverty has many dimensions, and poor communities in Ghana, are characterised by low income, malnutrition, ill health, illiteracy, and insecurity. The living standards surveys also reported that these different characteristics interact and combine to keep households, and at times the entire communities in persistent poverty. The dimensions of poverty usually considered by the Living Standard Surveys reports are consumption poverty, lack of access to services and limited human development which clearly show the intention to deal with the complex nature with which the term ‘poverty’ is used, and by the many different indicators proposed to monitor poverty.

Ghana's poverty rate uses specific consumption poverty line, reflecting the economic and social circumstances at specific times. The poverty line is the level of standard of living measure at which minimum consumption requirements can be met. In other words, the poverty line is the minimum consumption or income level necessary to meet basic needs of a person or a household. The national poverty rate is the proportion of the population living below the national poverty line. The Ghana Living Standards Surveys (GLSS) set two poverty lines, the extreme poverty line and the general poverty line.

The extreme poverty line focuses on what is needed to meet the nutritional requirements of a household. That is, individuals or households whose total expenditure fall below the minimum amount needed to meet only nutritional requirements are considered extremely poor. Even if these households allocate their entire budgets to only food, they would not be able to meet their minimum nutritional requirements, assuming they consume the average consumption basket.

On the other hand, the general or upper poverty line includes both food and non-food consumption. In other words, individuals or households consuming at levels above what is needed to meet only the nutritional requirements, and are able to meet some basic non-food needs fall on or under the upper poverty category. Other useful indicators of measuring poverty include the poverty gap ratio, which takes into account the distance of poor people from the poverty line, and the squared-poverty gap, which considers the degree of income inequality among poor people.

According to the 2010 Population and Housing Census Report on Non-Monetary Poverty, poverty is generally defined and measured based on consumption or income levels. However, poverty is widely recognized as multi-dimensional, and as a result, its definition and measurement based on only income or expenditure does not provide the true command of resources by individuals or households – neglecting benefits from help provided by family and friends, as well as consumption of public services such as education, health and housing. It is based on such criticisms and limitations that non-monetary poverty measurements have received attention as complementary measurements to income or consumption poverty measurements.

This research aims at measuring consumption poverty from non-consumption expenditure household data by using characteristics of households that are identified to be correlated with household consumption expenditure. The data set used for the study is the sixth round of the Ghana Living Standards Survey (2012/13) and the Multiple Indicators Cluster Survey (2011) conducted by the Ghana Statistical Service.

1.2 PROBLEM STATEMENT

The six series of Ghana Living Standards Surveys (GLSS) conducted by the Statistical Service have shown a general decline in poverty over the last three decades. However, poverty estimates are only measured when the next GLSS data is available. In-between the living standards surveys, no poverty estimates are available, and planning and policy are based on the recent past GLSS estimates. Moreover, the household expenditure data, which is used in estimating poverty indicators, becomes available after every seven years. This has reduced the frequency of measuring poverty indicators. Furthermore, much research on applying scientific methods to the available expenditure and non-expenditure household data to estimate poverty indicators have not been identified yet.

Thus, this has necessitated measurement of poverty indicators from other household characteristics, which are empirically recognized to have significant influence on poverty using multidimensional approach. These household characteristics are easy to collect since they require comparatively less time, and can also be included in other surveys which are conducted in-between GLSS.

This study seeks to measure poverty using household characteristics which are generally accepted as having significant influence on poverty in another household survey (Multiple Indicators Cluster Survey round four) and the most recent Ghana Living Standards Survey data from the Ghana Statistical Service by applying Multiple Imputation method.

1.3 OBJECTIVES OF THE STUDY

The main objective of the study is to provide a statistical model to estimate poverty incidence using household characteristics from a non-expenditure households' survey data, and applying multiple imputation regression technique with an expenditure survey as a baseline. The specific objectives of the study are:

i. To provide an alternative method of measuring poverty incidence to increase the frequency of poverty indicators measurement.

ii. To develop a non-consumption expenditure poverty model, that can be used to predict poverty incidence in-between the Ghana Living Standards Surveys.

1.4 RESEARCH QUESTIONS

The research questions are:

1. Is the proposed model developed from the consumption expenditure data (i.e. the Ghana Living Standard Survey) able to estimates poverty incidence from a non-monetary data set with minimum allowable error?
2. Is there a significant difference between the estimates obtained from the proposed model and that of the consumption expenditure estimates of poverty incidence from the Ghana Living Standards Survey?
3. How consistent and robust will the proposed model be in estimating poverty incidence using other surveys in-between Ghana Living Standards Surveys?

1.5 SIGNIFICANCE OF THE STUDY

Generally, it has been identified that the concepts of poverty have profoundly changed from the mere consideration of income or consumption (expenditure), to include dimensions such as deprivation and well-being. However, the measurements of poverty usually considered by the Living Standard Surveys are the consumption expenditure poverty. Gathering household consumption expenditure data is very expensive. It also requires a long period of time (12 months) and considerable effort. A number of conventional measures relating to household income and expenditure pattern have been used in estimating poverty levels. However income and expenditure dynamics are difficult to measure because reliable sources of data are hard to come by (Meng & Gregory, 2007).

This research will determine poverty estimates from other household characteristics which are widely recognized to have significant influence on poverty. These household characteristics are easier to collect or to include in other surveys. It is comparatively less expensive to collect, and does not require that long period of time to collect. As a result, policy makers would not be waiting for, between five and six years of the next living standards survey in order to obtain poverty estimates for monitoring, evaluation and decision making.

This research, using an alternative method of determining poverty estimates, will come out with a model, which will be consistent in predicting poverty estimates, and also could be referenced for further improvements in future studies.

1.6 ORGANISATION OF THE STUDY

This research is organized into five chapters. Chapter one consists of introduction, background of the study, problem statement, objectives of the study, research questions, significance of the study, scope and limitations. Chapter two covers literature review. Chapter three contains the methodology, which deals with source of data, sampling, data, stepwise regression technique, model assumptions and test of the assumptions, model validation and multiple imputation regression technique and statistical software used for the analysis. Chapter four comprises of the preliminary analysis, model building and its application to MICS4 data and results. Discussions, conclusion and recommendations are covered in chapter five.

1.7 LIMITATIONS

This study uses two data sets, GLSS6 and MICS4. The model is developed using data sets from GLSS6, which was conducted in 2012/2013 and MICS4 conducted in 2011. The data set of the most recent nationally representative household survey (GDHS 2014) is not available for this study, though the study intended to use it. Moreover, financial and other resources are also not available to personally conduct a nationally representative survey purposely for this study. Furthermore, some variables in GLSS6 data which are empirically proven to have significant influence on poverty (i.e. employment and "susu" variables) were not available in the MICS4 data set. The health variables identified to be common to both data sets had too many missing values in the GLSS6 data set.

CHAPTER TWO

LITERATURE REVIEW

2.0 INTRODUCTION

It is widely acknowledged that data availability is critical in the measurement and monitoring of poverty levels so as to make right decision and formulate and implement the required policies in the fight against poverty. Alkire (2014), asserted that poverty data from household surveys has increased in both quantity and frequency over the past 30 years, but still lags behind as compared to the data available on most other economic phenomena. The writer further states that the post-2015 agenda, Sustainable Development Goals (SDGs), identifies the need for regularly updated data to monitor the progress towards these goals.

In Ghana, existing literature on poverty mostly rely on the series of the Ghana Living Standards Surveys, and are mostly at the national level or to some extent on urban and rural localities. Each of the six rounds of these household expenditure surveys have field data collection alone covering one year and processing as well as other final activities also taking about the same time. Ultimately, the time interval between these household surveys data availability is seven years (Ghana Statistical Service, 2014).

This study tries to extend the literature by developing a predictive model with an expenditure survey and determining poverty incidence with a non-expenditure survey using the predictive model and multiple imputation regression techniques to obtain estimates of poverty statistics for the years that household expenditure surveys are not available. This is to frequently make available accurate and reliable poverty statistics at national, regional and urban-rural levels.

2.1 GENERAL DATA SOURCES FOR POVERTY ESTIMATES

Monitoring poverty requires comparisons of poverty profile across time and place. Appropriate comparisons is valid and relevant only if the data used to estimate the poverty statistics follow the same or a standardised methodology over time and across countries.

Over the years, national poverty estimates have been used to inform policy discussions on poverty alleviation in many developing countries. These poverty estimates have mainly relied on cross-sectional household expenditure surveys, which provide detailed poverty profile, and track incidence of poverty in various segments of the population. These cross-sectional surveys provide the raw data for most of the poverty assessments that are recently carried out regularly for many developing economies. Alkire (2014), argued that it is generally acknowledged that data on poverty are limited both in terms of frequency, coverage and content.

According to Alkire (2014), the limitation of data on poverty with regards to frequency is especially striking when compared with data availability concerning other economic variables such as Gross Domestic Product (GDP), inflation and Gross National Income (GNI). The author argued that GNI data is published annually, GDP and external debt statistics are available on a quarterly basis whereas poverty headcount data is available only when Living Standard Surveys are conducted usually between three and seven years. Again, inflation data is published monthly and stock market data is released every day.

In terms of coverage, as in most developing countries, the household expenditure surveys, which have been the main source of data for estimating poverty indicators, do not provide data for the lowest units of administration by the sample design since it is expensive and takes longer time, usually not less than one year to collect the data.

In order to expand coverage for poverty statistics, a further statistical methodology, such as application of Small Area Estimation (SAE) by Elbers, Lanjouw and Lanjouw (2003), have become appropriate for many developing countries to provide poverty estimates (or poverty maps) at the lower units of administration. The methodology developed by Elbers et al (2003) allow accurate estimates of consumption-based poverty and inequality at lower levels of disaggregation by combining information from censuses and household consumption surveys.

Daniels (2011), examined poverty trends across Uganda from 1995 to 2010 and used non-monetary indicators based on household assets, housing characteristics, and household size and composition. In a variation on poverty mapping methods, the author selected household characteristics that were available in four national "Demographic and Health Surveys" (DHS) and the 2005 Uganda National Household Survey (UNHS), an expenditure survey.

In terms of content, poverty data continues to miss information on important dimensions of poverty such as domestic violence and empowerment. Information on basic variables such as health remains quite limited in poverty data (Alkire, 2014).

The 2014 MDG report draws attention to the fact that reliable statistics for monitoring development remain inadequate in many countries. Data gaps, data quality, compliance with methodological standards and non-availability of disaggregated data are among the major challenges to MDG monitoring. Also the United Nations Statistics Division (UNSD), 2005, asserts that data from both developed and developing countries often fall short of the needs of researchers and policymakers who are interested in poverty issues.

It is observed (UNSD, 2005), that almost all countries have either a Household Income and Expenditure Survey (HIES) or a Household Budget Survey (HBS) for poverty assessment. However, methods used to measure consumption expenditures in these surveys vary widely, in terms of data collection (recall, family diaries, and individual diaries), reference periods over which consumption is observed, and whether households are observed only once or revisited during a year.

The UNSD (2005), argued that the major problem with these surveys is the short period over which consumption is observed. This is because respondents find it difficult to remember spending on frequent purchases but HIES and HBS typically use a very short reference period (e.g. a one-week recall or a two-week diary), which is typical of the household's usual standard of living.

On the other hand, poverty measurement requires accurate estimates of long-run welfare for each household. Nonetheless, many surveys that provided such long-run measures report expenditures and poverty on an annual basis by simply observing households for a week, fortnight, or month, with consumption from these periods respectively annualised by multiplying by 52, 26, or 12.

The problem with these annualised estimates and also with estimates that are collected and reported for shorter periods like a fortnight or a month is that random shocks, which occur during the observation period and are subsequently evened out over the rest of the year, get included along with the genuine between-household inequality in annual expenditures. Consequently, estimates of annual inequality are overstated resulting from setting the poverty line below the modal value of per capita expenditure, leading to an overstatement of the poverty head-count and other measures of poverty.

The UNSD (2005), stated that household-based surveys, both cross-sectional and panel or longitudinal provide important economic and social information about the human condition. It continues that very important characteristics of a household survey is whether the data are collected from the same households and individuals over time (called panel data) or if the data are collected from different households each time the survey is conducted (known as a repeated cross-sectional survey). It argued that generally, panel data provide much more information on poverty dynamics than repeated cross-sectional data, but panel data are somewhat more complicated to collect.

According to Chaudhuri, Jalan and Suryahadi (2002), cross-sectional household surveys are much more widely available than are panel household surveys. In Ghana, poverty assessment has mainly drawn on cross-sectional household data, and each of these cross-sectional data has different sample size based on the population sizes at the time.

Glewwe (2001), disputed the general consensus that household income and expenditure surveys or household budget surveys are the only sources of information for the purpose of measuring poverty. The writer suggested that, apart from income and expenditures surveys, other surveys such as labour force surveys and demographic and health surveys provide adequate information to understand the nature of poverty. Also UNSD (2005), proposed that several different types of household survey can be used to measure and analyse poverty, though very few of these surveys have poverty measurement as their primary objective. Thus statistical agencies have to carefully evaluate whether surveys that have other (or multiple) objectives can provide reliable data for measuring poverty.

2.2 DEFINITION OF POVERTY

Poverty has rich vocabulary, in all cultures and throughout history. The terms used to describe poverty include income or consumption poverty, human (under) development, social exclusion, ill-being, vulnerability, livelihood unsustainability, lack of capability and functioning, lack of basic needs, and relative deprivation. It is very vital to be precise in defining the term ‘poverty’ in order to measure it accurately for effective policy formulation and implementation. Poverty is a social science and moral concept, and many problems encountered in measuring poverty emanate because both concepts are most often confused.

The two basic theory of poverty in social science are ‘absolute poverty’ and ‘relative poverty’. Todaro and Smith (2012), defined absolute poverty as a situation of being unable to meet the minimum level of income, food, clothing, healthcare, shelter, and other essentials, and summarise the definition of relative poverty as the lack of collateral. Explicitly, relative poverty is when low-income individuals, whether one is absolutely poor or not, cannot borrow money, and generally cannot adequately educate one’s children or start and expand a business. These definitions clearly indicate that an individual can be relatively poor but not absolutely poor but once an individual is absolutely poor, inevitably, that person becomes relatively poor.

The World Bank reports that poverty has many dimensions, and that absolute poverty, which is the proportion of the population below national poverty line, measures poverty by the level of income/consumption available to an individual. It further states that a person is considered poor if his or her consumption or income level falls below some minimum level necessary to meet basic needs. This minimum level is usually called the "poverty line". It continues that poverty line varies across time and societies. Each country uses a line, which is appropriate to its level of development, societal norms and values (World Bank, 2000). The World Bank reiterates that national poverty lines are set to reflect a country’s specific economic and social circumstances, and therefore, are not intended for comparison across countries. It states further that local poverty lines tend to have higher purchasing power in rich countries, where more generous standards are used, than in poor countries (World Bank, 2007).

Furthermore, it defines national poverty rate as a "headcount" measure, which is by far the most commonly calculated measure of poverty. However, poverty rate fails to reflect the fact that among the poor there may be wide differences in income levels, with some located just below the national poverty line while others would be experiencing far greater shortfalls.

National poverty line is country specific, and cannot be used for international comparison. However, there is an international poverty line, which does not depend on the national poverty line. According to Todaro and Smith (2012), the international poverty line knows no national boundaries, it is independent of the level of national per capita income, and takes into account differing price levels by measuring poverty as anyone living on less than $1.25 a day or $2 per day in Purchasing Power Parity (PPP) dollars.

Townsend (2002), defined poverty as a situation where resources seriously fall below those commanded by the average individual or household to the extent that the poor are, in effect, excluded from ordinary living patterns, customs and activities. As resources for any individual or households are diminished, there exist a point where one suddenly withdraws from participation in the customs and activities demanded by culture. The point at which withdrawal escalates unduly to falling resources defines the poverty line or threshold.

Mack and Lansley's (1985), subsequent work on Poor Britain survey, which built on the work of Townsend, defined poverty as an 'enforced lack of socially perceived necessities' and sought to establish the public's view on 'what it is that people need for living in Britain in the 1980s'.

The World Bank (2005), alternatively defines poverty as relative deprivation, for example, as half mean income, or as exclusion from participation in society. According to Maxwell (1999), this definition started in France in the late 1970s, but has spread widely. The focus on multiple deprivation did not only include low income, poor housing, poor access to education and health, but also the process by which multiple deprivation occurs. The key areas for exclusion include democratic and legal systems, markets, state welfare provisions, and family and community: rights, resources and relationships.

Now, the underlining question asks whether there is a single comprehensive definition of poverty. The answer remains certainly not affirmative. However, it can be perceived that current viewpoints allow some simplification from the two major tightly linked concepts of relative poverty and relative deprivation. A careful consideration of the complex matrix of terminologies used in defining poverty is necessary for a precise definition and measurement of poverty.

2.3 MEASUREMENT OF POVERTY

There is no ideal measure of well-being, and analysts need to be aware of the strengths and limitations of any measure they use (World Bank, 2005). The World

Bank points out that measuring poverty require using information on welfare from household survey data and defining an indicator of welfare such as income or consumption per capita.

A major strength in the measurement of poverty is said to be the combination of both economic and social factors. The combination of economic and social indicators is likely to provide a better measure due to its explicit recognition that there is much more to welfare than economics (GSS, 2013). Though poverty studies embrace three broadly constructed definitional and measurement approaches – economic well-being, capability, and social exclusion – significant efforts are yet to be made to integrate them. The complex nature of defining what poverty really is indicates that the reductionist approach to poverty definition with too much emphasis on one aspect cannot give exhaustive understanding of the main poverty issue (Nolan and Whelan 2010).

There are a number of conceptual approaches to the measurement of well-being. The most common approach is to measure economic welfare based on household consumption expenditure or income. Certainly, there are also non-monetary measures of individual welfare, which include indicators such as infant mortality rates, life expectancy, the proportion of spending devoted to food, housing conditions, and child schooling. Well-being is a broader concept than economic welfare, which only measures a person's command over commodities (World Bank, 2005).

According to Osberg and Sharpe (2005), distributional issues, particularly poverty and social exclusion should not be considered in isolation, as if tradeoffs between them might not matter. However, to be able to generate adequate index of economic and social indicators that provides a good and bigger picture of the situation of the poor, a better measure of access to resources needed for a decent standard of living is required (GSS, 2013).

Material deprivation (e.g. poor housing) may be distinguished from social deprivation (e.g. lack of rights or of power), while patterns of deprivation may manifest inconsistencies as, for instance, when people who are materially prosperous are deprived in their work situation – or vice versa (GSS, 2013). According to Townsend (2010), people are rich or poor according to their share of the resources that are available to all. The writer further stated that, the general theory, then, should be that individuals and families whose resources over time fall short of the resources commanded by the average individual or family in the community in which they live are in poverty, whether that community is a local, national or international one.

The extent to which poverty is perceived as a problem is influenced by the way in which it is measured, and its measurement is also predisposed to what constitutes poverty. It must be acknowledged that poverty is a dynamic concept, and that there are gaps in the existing knowledge of the qualitative and quantitative features of human needs (GSS, 2013).

It is evident that poverty has been defined in many ways, with each definition depending on how poverty was conceptualized which has led to a particular identification of the poor. However, Maliki (2011) pointed out that poverty level can be measured generally on the basis of two approaches: the monetary (material or utilitarian) and non-monetary (non-material or non-utilitarian). The monetary, capability and non-monetary approaches in the measurement of poverty are considered in the subsequent sections.

2.3.1 Monetary Approach to Poverty Measurement

This approach defines poverty with a shortfall in monetary income or consumption from a fixed poverty line. If poverty is measured based on household consumption or expenditure per capita then it is helpful to use an expenditure function, which shows the minimum expense required to meet a given level of utility *u*, which is derived from a vector of goods *x*, at prices *p* (World Bank, 2005).

Measuring 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 *u*, which is derived from a vector of goods *x*, at prices *p*. It can be derived 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 (World Bank, 2005).

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

$y_i = \boldsymbol{p} * \boldsymbol{q} = e\,(\boldsymbol{p}, \boldsymbol{x}, u)$

where $\boldsymbol{p}$ = a vector of prices of goods and services, $\boldsymbol{q}$ = a vector of quantities of goods and services consumed, *e* (.) is an expenditure function, $\boldsymbol{x}$ = a vector of household characteristics (e.g. number of adults, number of young children, etc.) and u = the level of "utility" or well-being achieved by the household. In other words, given the prices (p) that it faces, and its demographic characteristics (x), y_i measures the spending that is needed to reach utility level u.

The actual level of y_i is computed from household survey data that include information on consumption. As y_i is computed, per capita household consumption for every individual in the household is also constructed. This implicitly assumes that consumption is shared equally among household members. This approach assumes that all individuals in the household have the same needs. This is a strong assumption, since in reality; individuals are unique and have different needs based on their individual characteristics (age, gender, job, etc.). There are several factors that complicate the estimation per capita consumption.

Traditionally, household welfare is measured in monetary terms. The two most obvious candidates of monetary measurement of household welfare are income and expenditure. Most rich countries measure poverty using income, while most poor countries use expenditure. The logic behind this is that in rich countries, income is comparatively easy to measure (much of it comes from wages and salaries) while expenditure is complex and hard to quantify. In contrast, in less developed countries, income is difficult to measure (much of it comes from self- employment), while expenditure is simple and hence easier to estimate.

2.3.2 Capability Approach to Poverty Measurement

This approach focuses on failure of some basic capability of functioning (Sen, 1985). The capability approach places emphasis on what people are able to do and be, as opposed to what they have, or how they feel. Sen argues that, in analysing well-being, we should shift our focus from 'the means of living', such as income, to the 'actual opportunities a person has', namely their functioning and capabilities (Sen, 2000). 'Functioning' refer to the various things a person succeeds in 'doing or being', such as participating in the life of society, being healthy, and so forth, while 'capabilities' refer to a person's real or substantive freedom to achieve such functioning; for example, the ability to take part in the life of society (Sen, 1993). Most important is the emphasis on real or substantive – as opposed to formal – freedom, since capabilities are opportunities that one could exercise if so desired. The capability approach places particular emphasis on the capabilities a person has, irrespective of whether they choose to exercise these or not.

The capability approach to poverty measurement presents three main operational issues - definition of basic capabilities, measurement of these capabilities and aggregation. Methods of defining basic capabilities invariably amount to the establishment of a list of sensitive basic needs using various fundamental criteria. Most of these techniques have led to similar interpretation of minimal essential capabilities as being constituted by health, nutrition and education indicators. In practice, these basic capabilities are measured through designing functions of life expectancy, morbidity and nutrition levels.

The capability approach asks the central role often given to income in poverty measurement. Sen (2000), distinguishes between the actual opportunities, or capabilities, a person has, which he argues are intrinsically important, and their income, which is merely a means to such opportunities, and whose importance is thus both instrumental and contingent (Sen, 2000). This relates to the distinction between direct and indirect concepts of poverty drawn by Ringen (1988). Direct concepts of poverty focus on cases where living standards fall below a certain minimum level, and typically assume that this is because of a lack of resources. Indicates that indirect concepts focus on cases where resources fall below a certain point, and assume that this results in a low standard of living (Berthoud and Bryan, 2011).

Certainly, such distinctions would be of little importance if low income were a good proxy for deprivation. However, the capability approach is of the view that this is not likely to be the situation in that people have varying needs, and therefore, will require different levels of resources in order to achieve the same standard of living.

While the asset index approach of Filmer and Pritchett (2001) has not been used to directly study poverty, a related method has been developed by Sahn and Stifel (2000) to make poverty comparisons across time and space for 11 African countries. In this method, Demographic and Health Survey (DHS) data from all 11 countries are pooled and an asset index is formed using the method of factor analysis. Unlike the method of principal components, which uses all the variability in an item, factor analysis allows some variability to be unique, with only the variability that is common with the other items used to form the asset index.

2.3.3 Multi-dimensional Approach to Poverty Measurement

In 2010, the UNDP introduced the Multi-dimensional Poverty Index (MPI), as an index to capture the multifaceted and multi-dimensional nature of poverty. The MPI is grounded in Sen's (1985, 1993) capability approach, and attempts to capture the multiple deprivations that plague households. According to the UNDP (2010), the MPI replaces the HDI, which has been published since the 1990s. It adds that the MPI addresses a key limitation of the HDI as it captures how many people experience overlapping deprivations and how many deprivations they face on average. The MPI can be disaggregated by region, ethnicity and other groupings as well as by dimension (health, education and standard of living), thus making it an appropriate tool for policy-makers (GSS, 2013).

According to (GSS, 2013), Multidimensional poverty measurement relate to the capability approach insofar as they provide information by virtue of which it may be possible to be more accurate in reducing people's capability deprivations (Sen, 1985). GSS continues that there is increasing consensus that poverty is now an intrinsically multidimensional phenomenon following Sen's (1985, 1993, and 2000) pioneering capability approach. This has resulted in many researchers proposing different multidimensional poverty measures. Departing from this, Alkire and Foster (2007) proposed a new family of multi-dimensional poverty measures, which is a variant of the extensively used Foster, Greer and Thorbecke's (1984) class of one-dimension poverty measures, simply referred to as FGT. The dimension adjusted FGT measures keep the simple structure of the one-dimension case and satisfy a set of convenient properties, among which disaggregation across population subgroups and the possibility to break it down by dimension are useful for policy purposes.

2.4 PATTERNS AND TRENDS OF POVERTY IN GHANA

Poor communities in Ghana suffer low income, malnutrition, ill health, illiteracy and insecurity. They also express a sense of powerlessness and isolation. These different aspects of poverty interact to keep the standard of living of the poor below some minimum level required to meet basic needs. This section examines poverty patterns and trends from consumption poverty, household assets, and human development. Ghana Statistical Service is the sole agency which collects data on welfare (i.e. the Ghana Living Standards Survey) and reports on trends and patterns of poverty. Other researchers draw on these data sets for further analysis and findings.

2.4.1 Consumption Poverty Trends

Over the years, Ghana's poverty analysis has focused on consumption poverty, identifying the poor as those who lack command over basic consumption needs. In addition to the poverty line, an extreme poverty line, which is, resources required for a minimum food consumption basket that can provide adequate calories to a household, is also estimated. There are two aspects of consumption poverty that are mainly of interest, incidence of poverty and depth of poverty. Thus, the various poverty indices available are combinations of one or both of these aspects. Incidence of poverty in looks at the proportion of the Ghanaian population identified as living below the national absolute poverty line, while depth of poverty focuses on the extent to which those defined as poor fall below the poverty line.

The poverty reduction programmes including the Economic Recovery Programme (initiated in 1983), the Structural Adjustment Programme, Ghana Poverty Reduction Strategy (GPRS I, 2001-2004), Growth and Poverty Reduction Strategy (GPRS II, 2005-2008) and the Ghana Shared Growth and Development Agenda (GSGDA I & II, 2009-Date) implemented to salvage the economy from deteriorating and also to improve the living conditions of the poor has resulted in the general trend of poverty incidence in Ghana falling continuously since 1990. According to the Ghana Statistical Service (2000), with an absolute poverty line of GH¢ 90.00 per annum, the percentage of the Ghanaian population defined as poor fell from almost 51.7 percent in 1991/92 to 39.5 percent in 1998/99. The decline, however, was concentrated in Accra and the forest localities. The remaining localities, both urban and rural, saw very modest reduction in poverty incidence whereas the urban savannah had increased proportion of the population of the poor.

During this period of poverty reduction (i.e. 1990 to 1992), the GDP growth rate remained at about five percent per annum and this accompanied by significant improvement in key social indicators. Infant mortality rate dropped from 77 to 66 per 1000 live births, child mortality rate also decreased from 84 to 57 per 1000 live births, malnutrition rate declined from 31 to 26 percent and total fertility rate also reduced from 6.4 to 5.5 (World Bank, 1995).

The fifth round of the Ghana Living Standards Survey (2005/06) with an absolute poverty line of GH¢ 370.89 revealed a further decline in the incidence of poverty in Ghana from 39.5 percent in 1998/99 to 28.5 percent in 2005/06. This decline, unlike the previous, was reflected significantly in all localities except Greater Accra which experienced an increase (GSS, 2007). The recent GLSS6 adjusted household consumption to reflect the rebasing of the Consumer Price Index (CPI) in 2012 reported a further reduction in the proportion of the population identified as poor, with an absolute poverty line of GH¢ 1,314.00. The incidence of poverty has reduced from 28.5 percent in 2005/06 to 24.2 percent in 2012/13 with a poverty gap index of 7.8 percent. In other words, the mean income of the poor falls below the poverty line by 7.8 percent. These percentages indicate that about 6.4 million people currently in Ghana are poor (GSS, 2014).

2.4.2 Household Assets Ownership Trends

Household assets included key consumer durable goods and changes in household ownership of such assets were considered as an indicator of changing living standards of households. Indeed, availability of electricity significantly determined a household's ownership of certain assets. However, the household asset ownership can be used as a proxy indicator of living standards. Information on the proportion of households owning different key consumer durable goods between 1991/92 and 2012/13 are presented in Table2.1.

Ghana Statistical Service (2000) reported a large increase in the proportion households that owned household assets in both rural and urban localities between 1991/92 and 1998/99. Noticeably among the assets were refrigerators, radios, televisions, fans, electric irons and bicycles. Besides bicycles which are mostly owned by households in the savannah areas in Ghana, much higher proportions of households in urban than in rural areas owned these assets. Most likely, this reflects not just higher incomes in urban areas but also supply factors including wider access to electricity as a result of the rural electrification programme which was ongoing.

The pattern of changes in ownership of household assets between 1998/99 and 2005/06 is different for urban and rural households. While the increased in ownership of items were relatively significant for only three assets in rural areas, increased in urban areas were reflected in six items. Ownership of radio sets and mobile phones show large increases in both rural and urban areas, but video recorders, television sets and cooking stoves show significant increases in their ownership in urban areas. In addition to higher incomes in urban areas, supply factors including broader access to electricity and liquified petroleum gas (LPG) contributed to the increased asset ownership (GSS, 2007).

The proportions of households with ownership of most of the durable goods covered in the GLSS6 showed further increases in 2012/13. These increases were observed in both urban and rural areas but often were higher for wealthier groups, with greater disparity among urban households. Ownership of durable goods remains much lower in rural than urban localities, even among households of similar overall living standards (GSS, 2014).

Increased ownership of household assets over the years corresponds with the continuous fall in the consumption based incidence of poverty indicating that more households may have been alleviated from poverty. The increased asset ownership also partly indicates that more households are not just able to meet the minimum food requirement but also have the ability to own at least one household asset.

Table 2.1: Proportion of households with ownership of assets, by locality of residence

Asset	1991/92			1998/99			2005/06			2012/13		
	Rural	Urban	Total	Rural	Urban	Total	Rural	Urban	Total	Rural	Urban	Total
Sewing machine	22.6	35.8	27.2	28.6	33.1	30.3	18.4	24.5	21.0	13.8	17.5	15.8
Stove	6.5	27.1	13.7	6.7	23.2	12.8	6.9	32.6	18.0	10.0	46.4	30.2
Refrigerator	1.8	20.3	8.2	7.3	32.8	16.6	7.7	38.8	21.2	15.0	52.9	36.0
Fan	3.9	33.7	14.3	11.3	44.9	23.6	12.3	53.8	30.2	25.2	69.0	49.5
Radio	34.7	54.3	41.5	47.2	65.2	53.8	70.8	77.3	73.6	67.1	64.8	65.8
TV	0.1	2.3	0.9	1.5	8.7	4.1	14.5	53.1	31.2	34.4	75.4	57.1
Camera	3.2	25.2	10.9	12.1	40.1	22.4	1.5	4.6	2.9	0.7	3.9	2.5
Mobile phone	*	*	*	*	*	*	6.4	35.7	19.1	70.3	88.3	80.2
Computer	1.2	3.7	2.0	1.4	5.1	2.7	0.5	4.2	2.1	4.2	17.6	11.6
Bicycle	4.0	37.6	15.7	11.0	46.1	23.8	29.1	13.7	22.5	28.1	13.8	20.2
Motorcycle	19.1	8.7	15.5	23.5	11.8	19.2	2.4	2.4	2.4	8.9	6.2	7.4
Car	0.7	4.0	1.9	1.3	5.0	2.6	1.2	5.4	3.0	2.4	7.0	4.9

Source: Computed from the Ghana Living Standards Survey, 1991/1992 - 2012/13

** Not included in the basket of household consumption at the time.*

2.4.3 Human Development Trends

Human development in terms of education and health were considered “basic needs” and complementary to the consumption-based welfare indicator. Education and Health possess characteristics as public goods and are, therefore, conceptually difficult to measure in monetary terms. According to the World Health Organization (WHO, 1948), health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity. Education has been identified as an important tool in providing people with basic knowledge, skills as well as competencies to improve their standard of living and quality of life.

In the period 1998-1999, Ghanaians were less likely to consult well-qualified health personnel, or to go to a hospital when they are ill or injured. Increasing numbers did not consult anyone at all. The survey period from 2005/06 to 2012/13 saw increased rates of access to a range of health services. However, disparities remain between urban and rural areas. Compared to 1998/99 and 2005/06, individuals were more likely now to consult doctors and visit health facilities in 2005/06. Consultation with pharmacists or chemical sellers when ill or injured decreased in 1998/99. The percentage of individuals ill or injured who did not consult any health practitioner also declined over the years. This pattern was observed in all income groups in both rural and urban areas.

In terms of education, school attendance rates in primary, JHS and SHS improved over this period. The savannah areas reported the lowest school attendance rates. The increased net school attendance rates at the JHS level were much higher for girls than boys. The reverse was the situation for SHS. Even with these increases, net attendance rate at SHS were much lower than at the primary and JHS levels, and especially so in rural areas.

Ghana has made huge amount of progress in alleviating poverty. More importantly, the country has met the first MDG target of halving poverty between 1990 and 2015, reducing it from 51.7 percent of the population in 1992 to 24.2 percent in 2013. Progress has also been made in many important areas such as education, healthcare and infrastructure, which clearly reflects in the gains made in reducing poverty among different segments of the population. Generally, poverty remains a rural phenomenon, with those in rural Savannah being mostly affected.

2.5 MULTIPLE IMPUTATION TECHNIQUES

There are many ways to handle incomplete data. Complete Case Analysis (CCA) and Available Case Analysis (ACA, pairwise deletion) are probably the most common. Schafer and Graham, (2002) argue that removing all incomplete cases and relying on cases with complete records, as is done with CCA and ACA, is simple but yields loss of data. It also draws on some simple assumptions to produce unbiased estimates. Schafer and Graham further suggest that one alternative approach to handling missing values is to use imputation which does not lose data and makes flawless assumptions. Little and Rubin (1987) point out that, using Single Imputation (SI), where each missing observation is replaced by an imputed value to create a "complete" data set fails to account for the uncertainty about the missing values, which usually cause standard errors to be too small and p-values to be too low. Schafer and Graham (2002) agree that Multiple Imputation (MI) introduced by Rubin (1987), is a generally accepted method to handle incomplete data.

Multiple imputation is a general approach to the problem of missing data that is available in several commonly used statistical packages. It allows for the uncertainty about the missing data by creating several different plausible imputed data sets and appropriately combines results obtained from each of the imputed data set. It was mainly thought of as a way to handle nonresponse in a complex survey context when it was first suggested (Rubin, 1987). However, MI became accepted and useful in other settings as well within a short time (Rubin, 1996). Hornblad (2013) asserts that recently, the application of MI has grown and includes many statistical areas like clinical trials, epidemiology and longitudinal studies.

According to Rubin (1987), the general idea behind MI is really quite simple and straightforward. In the author's view, instead of imputing one value for each missing observation, *m* greater than or equal to 2 plausible values are imputed creating *m* "complete" data sets. The difference between the imputed data sets reflects the uncertainty caused by imputations. The MI procedure includes three steps. First, create the multiple imputed data sets, preferably by random draws from a posterior distribution. Second, analyze each data set separately. Third, combine the results into a single set of parameters, standard errors and test statistics using certain rules.

2.6 SUMMARY OF REVIEWED WORKS

The literature reviewed points out the two main methods of measuring poverty: economic approach and non-monetary approach. The economic measures of welfare are the expenditure approach and income approach. The non-monetary methods include capability approach, human development index, asset index and multi-dimensional poverty index. The most common approach identified from the existing literature is the economic measure of welfare based on household consumption expenditure or income.

From the literature, most rich countries measure poverty using income, while most poor countries use expenditure and the reason is that rich countries easily measure (much of it comes from wages and salaries) income while expenditure is complex and hard to measure. On the other hand, in less developed countries, income is difficult to measure (much of it comes from self- employment), while expenditure is easier to measure. The non-monetary measures of poverty are most often used as proxy indicators for the economic measure of welfare.

Ghana, belonging to the less developed countries, mainly uses the economic welfare measure of poverty based on household consumption expenditure. The MDG achievement made in halving poverty by 2015 resulted from tracking poverty incidence based on household consumption.

Empirically, missing values are the treated in two main ways: case deletion methods and imputations methods. The case deletion methods include Complete Case Analysis (CCA) and Available Case Analysis (ACA, pairwise deletion) which are the most commonly used. The imputation methods include the Single Imputation (SI) method and Multiple Imputation (MI) method. Due to the limitations associated with the SI method, as stated in section 2.5, the multiple imputation method was developed, which takes care of the problems with the SI; hence most researchers use multiple imputations to handle missing values.

Multiple imputations have been used by researchers mainly to impute handle missing values in a data set or to compare with other methods to test its performance. This thesis tries to apply multiple imputations to estimate incidence of poverty by imputing household expenditure values for households in a non-expenditure household survey data set. This thesis estimates a household expenditure variable, by multiple imputation technique, into a non-expenditure household survey data set to further estimate poverty incidence.

This study combined the economic approach and the non-monetary approach to estimate poverty incidence using multiple imputation technique. The economic approach formed the baseline, and enabled the estimation of household expenditure variable for the households in the non-monetary data set. The technique provided a statistical model to estimate poverty incidence from the estimated household expenditure. This provides an alternative method of measuring poverty incidence which will help increase the frequency of poverty indicators measurement.

CHAPTER THREE

METHODOLOGY

3.0 INTRODUCTION

This chapter provides a description of the data sets used for the analysis, explains the regression technique employed for the selection of the variables, and presents how the household expenditure for other non-expenditure survey (MICS4) is imputed. The data set of the sixth round of the Ghana Living Standards Survey 2012/2013 (GLSS6) and the fourth round of the Multiple Indictor Cluster Survey 2011 (MICS4) are used in this study. A stepwise regression method is used to develop the model using the GLSS6 data set. The final model is then used to impute household expenditure for the MICS4 data set using multiple imputation method. Finally, poverty indicators such as poverty headcounts are estimated from the imputed values.

3.1 DATA DESCRIPTION

The data sets used in this study, MICS 2011 and GLSS6, are household surveys carried out by the Ghana Statistical Service (GSS) in 2011 and 2012/13 respectively. As the names of the surveys indicate, GLSS6 focuses on various aspects of living conditions whereas MICS, which is an international household survey programme supported by UNICEF, collects reliable, disaggregated and internationally comparable statistics on households and health conditions of women (15-49 years), men (15-59 years) and children (under 5 years) in the households. The GLSS6 is the sixth round of the living conditions survey in Ghana while the MICS 2011 in Ghana, forms part of the 4^{th} round of MICS surveys (MICS4). Sections on demography, education, health and housing characteristics are common to both surveys. The study area for both surveys is the entire country, and the study population is all households in Ghana.

3.1.1 Sampling Design

The national sampling frame, prepared from the 2010 Population and Housing Census was the Primary Sampling Frame (PSF) used for the GLSS6 and MICS4. Again, the two surveys used the same sampling design by adopting a two-stage stratified random sampling design. The first stage selected Enumeration Areas (EAs) or Clusters from the national sampling frame. The second stage selected households from the sampled EAs. The random sampling of EAs and households were without replacement.

The selected EAs were allocated to the 10 regions according to Probability Proportional to Size of the region's population (PPS). Consequently, these EAs were divided into Urban and Rural EAs giving us a total of 20 strata. All households in each selected EAs were listed to obtain a Secondary Sampling Frame (SSF). Thus, the required number of households was systematically selected from the SSF.

3.1.2 Sample Size

The GLSS6 sampled 1,200 clusters from the national sampling frame and 15 households from each of the already sampled clusters yielding a total of 18,000 households.

However, 16,772 households were successfully interviewed yielding a response rate of 93.2 percent. In addition to the regions as the primary strata, three ecological zones; Savannah, Forest and Costal Zones, were used as another strata for analysis to inform policy, especially the Savannah Accelerated Development Agenda (SADA).

The field data collection lasted for 12 months in which these questionnaires were administered to the selected households in the sampled clusters. Table 3.1 shows the distribution of the clusters and households selected by region and urban-rural localities.

Table 3.1: Distribution of households sampled and interviewed by region and locality

Region	EAs/Clusters			Number of Households			Households interviewed
	Total	Urban	Rural	Total	Urban	Rural	
Ghana	1,200	545	655	18,000	8,175	9,825	16,772
Western	120	51	69	1,800	765	1,035	1,718
Central	116	55	61	1,740	825	915	1,602
Greater Accra	144	130	14	2,160	1,950	210	1,924
Volta	116	39	77	1,740	585	1,155	1,574
Eastern	128	56	72	1,920	840	1,080	1,804
Ashanti	148	90	58	2,220	1,350	870	1,981
Brong Ahafo	116	52	64	1,740	780	960	1,621
Northern	116	35	81	1,740	525	1,215	1,702
Upper East	100	21	79	1,500	315	1,185	1,447
Upper West	96	16	80	1,440	240	1,200	1,399

Source: GSS, Ghana Living Standard Survey Round Six (6), 2012/13

The MICS4 also sample 15 households from each of the selected 810 EAs giving us a total of 12,150 households sampled for the fourth round of the Multiple Indicator Cluster Survey. Table 3.2 provides the summary of distribution of clusters and households sampled and interviewed by region and locality.

Table 3.2: Distribution of households sampled and interviewed by region and locality

Region	Number of clusters			Number of households			Households interviewed
	Total	Urban	Rural	Total	Urban	Rural	
Ghana	810	309	501	12,150	4,635	7,515	11,925
Western	52	20	32	780	300	480	757
Central	134	56	78	2,010	840	1,170	1,989
Greater Accra	67	61	6	1,005	915	90	989
Volta	52	17	35	780	255	525	771
Eastern	52	22	30	780	330	450	767
Ashanti	67	37	30	1,005	555	450	993
Brong Ahafo	52	21	31	780	315	465	718
Northern	134	40	94	2,010	600	1,410	1,972
Upper East	100	19	81	1,500	285	1,215	1,475
Upper West	100	16	84	1,500	240	1,260	1,494

Source: GSS, Multiple Indicator Cluster Survey Round Four (4), 2011

3.1.3 Questionnaire and Data Collection

The GLSS6 used four separate questionnaires. These questionnaires were almost entirely pre-coded. The questionnaires were:

i. A household questionnaire which was divided into two parts, Part A and Part B. These were used to collect information on household characteristics including income and expenditure information of the household and its members.

ii. A community questionnaire was used to collect data on available services, economic activities, access to markets and social capital in the environments where the households reside;

iii. A price questionnaire was also used to collect price data to allow for cost of living adjustments; and

iv. Facility questionnaire was administered to local service providers to collect data on types and quality of service available to households.

The GLSS6 field data collection covered a period of one year, starting from October 2012 and ending in October 2013. Data was collected from the selected households in the sampled EAs.

The MICS4 also used four separate questionnaires each of which was almost completely pre-coded. The questionnaires used were:

i. Household Questionnaire: This was administered to gather information on household characteristics, facilities and assets.

ii. Woman Questionnaire: This was used to collect information on females 15-49 years identified in the household interviewed. The information included issues on reproductive and maternal health, HIV and AIDS and contraceptive use.

iii. Man Questionnaire: This was administered to males 15-59 years identified in the households interviewed. The issues covered in the questionnaire were similar to that of the woman questionnaire.

iv. Child Questionnaire: This questionnaire collected information on children below 5 years. Their mothers responded to the questions on their behalf. The issues covered in this questionnaire include nutrition, immunization and malaria.

The field data collection lasted for three months in which these questionnaires were administered to the selected households in the sampled EAs.

3.2 VARIABLE SELECTION AND MODELING

Although many variables can potentially predict welfare, it is important to select variables that closely relate to and explain a large part of the level of household welfare. These variables or indicators should be easy to measure and verifiable, and they should not be easy to hide or manipulate. According to the United Nations Statistics Division (2005) variables used as proxies of household welfare are usually the following:

i. **Location variables:** Urban/rural, region/district/province and ecological features.

ii. **Demographic characteristics:** Age, gender, marital status, ethnicity, education, and occupation of the household head, household size, dependency ratio and household literacy rate.

iii. **Dwelling characteristics:** Ownership and occupancy status of dwelling, number of rooms, material of wall, roof, and floor, type of kitchen, toilet, sewerage, and garbage disposal method, main source of lighting, cooking fuel, heating, and drinking water, distance to bus stop, roads, markets, schools, and health centers.

iv. **Ownership of productive and durable assets:** Ownership of livestock and poultry, agricultural land, TV, refrigerator, cooking stove, dishwasher, air conditioner, sewing machine, VCR/DVD player, computer, bicycle, motorcycle, car, mobile phone, land line, generator, iron, dish antenna, sofa, bed, closet and agricultural equipment.

3.2.1 Candidate Variable Selection

A two stage method is used to select the candidate variables for the analysis and modeling. At the first stage, the questionnaires of the two Surveys, GLSS6 and MICS4, are compared to identify questions with the same wording and response options on household characteristics. This was done manually to select the candidate variables. Thus, the data sets of these surveys were prepared to contain only those variables.

The second stage was where a stepwise regression procedure is applied to the candidate variables in the GLSS6 data set to select the variables that largely explain the variability in the household consumption expenditure and also obtain the parameter estimates of the model.

3.2.2 Stepwise Regression Model

A stepwise regression procedure selects the "best" subset of the predictor variables which relates significantly with the response variables sequentially. It is used in order to economize on computational efforts. Essentially, this method develops a sequence of regression models, adding or deleting a predictor variable at each step, with the criterion for adding or deleting a variable stated equivalently in terms of error sum of squares reduction, coefficient of partial correlation, t* statistic, or F* statistic and the p-value (Kutner, Nachtsheim, Neter and Li, 2005).

In this study, the t* statistic and their corresponding p-values are used for the usual tests of regression parameters. The stepwise regression model applied to the GLSS6 data set is of the form:

$Ln(y) = \beta' X + \mathcal{E}$ ………………………………………………………………... **(1)**

which is further expressed as

$lny_h = \alpha + \beta_1 x_{1h} + \beta_2 x_{2h} + \cdots + \beta_k x_{kh} + \varepsilon_h,$ ………………………………... **(2)**

where $h = 1, 2, \ldots, n$ is the number of households (or household heads).

$\boldsymbol{lny_h}$ = the natural log of household consumption expenditure per capita,

$\boldsymbol{\beta_i}$ ($i = 1, 2, \ldots, k$) = the regression coefficients of household characteristic i,

$\boldsymbol{x_{ih}}$ = the independent regressors of household characteristic i of household h, and

$\mathcal{E}_h$ = the error term assumed to be independent and identically normally distributed with a zero mean and variance Σ (i.e N (0, Σ)) for household *h*.

The Box-Cox transforms are a list of family of power transformations helps identifies most appropriate transformation of the response variable (*Y*). The Box-Cox transforms help to identify which transformation of *Y* is most appropriate for correcting skewness of the distributions of the error terms, unequal error variances, and nonlinearity of the regression function. The Box-Cox transforms are of the form $Y' = Y\mu$, where μ is a parameter to be determined from the data (Kutner et al, 2005). In this Study, the transformation which appropriately fits the data is $Y' = log_e Y$; where $\mu = 0$, hence the model (2) specified.

The stepwise regression routine fits a linear regression model for each of the *k-1* potential regressors. In this approach, the regressors are added to the model one at a time. At each step, the regressor with the largest t* statistic or smaller p-value is a candidate for addition. If this t* value exceeds a predetermined level, or if the corresponding p-value is less than a predetermined significance level of α, then the regressor is added. In other words, the regressor that produces significant t* statistic or p-value is added.

On the other hand, the regressor is removed if it does not produce a t* statistic that is significant. The regressors are re-examined to remove any one that entered the model earlier and has become not significant (or redundant) in the presence of recently added regressor. It is only after this check is made and the necessary removals accomplished can another regressor be added to the model (Kutner et al, 2005).

The stepwise process ends when none of the regressors outside the model has a t* statistic significant for entry and every regressor in the model is significant at the stay level, or when the regressor to be added to the model is the one just deleted from it.

3.2.3 Model Assumptions

The following assumptions stated underlie the regression model:

i. The error term ε is assumed to be independently and identically normally distributed with mean 0 and variance Σ. Symbolically, this is written as $\varepsilon \sim N(0, \Sigma)$ with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \Sigma = \sigma^2 I$ where I is a unit matrix of dimension n x k. The error term has constant variance (Rencher, 2002).

ii. The random variable $Y = (y_1, y_2, \ldots, y_k)$ is independent and normally distributed with mean $X\beta$ and variance $\sigma^2 I$. That is, $Y \sim N_n(X\beta, \sigma^2 I_{(n)})$. This implies that the distribution of Y has the same variance. (Rencher, 2002).

iii. The elements in the design matrix X are linearly independent. Also, the fitted model $\widehat{Y}$ is a linear function of the predictors X. (Kutner et al, 2005).

iv. The distribution of the estimated parameters $\hat{\beta}_i$ $(i = 1, 2, 3, \ldots, k)$ are normal with mean β and variance $\sigma^2 (X'X)^{-1}$. (Kutner et al, 2005).

3.2.4 Estimating Model Parameters

The method of Least Squares Estimation is employed in estimating the regression parameters using the stepwise regression technique. The Least Square Estimate of the parameters β minimizes the error Ɛ, and it is given by:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

The estimation of the parameter vector is obtained as follows:

Let Y = lny_h, which is the natural logarithm of the original response variable.

Given $Y = X\beta + \varepsilon$, we make Ɛ the subject and obtain $\varepsilon = Y - X\beta$

The Error Sum of Squares (SSE) can be obtained by the transpose of the matrix Ɛ by itself. That is, $SSE = \varepsilon^T\varepsilon$

Hence, $\varepsilon^T\varepsilon = (Y - X\beta)^T (Y - X\beta)$ ………………………………………..… **(3.1)**

Expanding the transposed bracket we obtain

$$\varepsilon^T\varepsilon = (Y^T - \beta^T X^T)(Y - X\beta)$$

Expanding the two brackets we have

$$\varepsilon^T\varepsilon = Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta \quad \text{since } (X\beta)^T = X^T\beta^T$$

Now, $\varepsilon^T\varepsilon = Y^T Y - \beta^T (X^T Y) - \beta^T(X^T Y) + \beta^T(X^T X)\beta$

Grouping like terms we get

$$\varepsilon^T\varepsilon = Y^T Y - 2\beta^T (X^T Y) + \beta^T(X^T X)\beta \text{ ……………………………..….. } \mathbf{(3.2)}$$

Now, to obtain minimum error, we minimize SSE by differentiating SSE with respect to β, and equate the derivative to zero. That is $\frac{d(SSE)}{d\beta} = \frac{d(\varepsilon^T\varepsilon)}{d\beta} = 0$

Now, differentiating equation (1) with respect to β gives;

$$\frac{d(SSE)}{d\beta} = \frac{d(\varepsilon^T\varepsilon)}{d\beta} = -2(X^T\mathrm{Y}) + 2(X^T\mathrm{X})\beta \quad \text{.....................................} \quad \textbf{(3.3)}$$

Taking the second derivative, we prove that the error is minimized when the outcome is positive. That is, $\frac{d^2(SSE)}{d\beta^2} = 2(X^T\mathrm{X}) > 0$

The second derivative is positive, indicating that a minimum is obtained.

Therefore, from (3), $\frac{d(SSE)}{d\beta} = 0$

This leads to the results below.

$$2(X^T\mathrm{X})\beta = 2(X^T\mathrm{Y})$$

$$\hat{\beta} = (X^T\mathrm{X})^{-1}(X^T\mathrm{Y})$$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

In this case, Y = $Ln(y)$, $\beta = (\beta_1, \beta_2, \beta_3 \dots \beta_k)$ *and* $\mathrm{Y} = (y_1, y_2, y_3, \dots, y_h)$.

Thus, the estimated $\hat{\beta}$ can be written as:

$$\hat{\beta} = (X'\mathrm{X})^{-1}(X'\mathrm{Y}) = (X'\mathrm{X})^{-1}X'(y_1, y_2, y_3, \dots y_h),$$

By expanding *Y* we get

$$\hat{\beta} = ((X'\mathrm{X})^{-1}X'y_1, (X'\mathrm{X})^{-1}X'y_2, (X'\mathrm{X})^{-1}X'y_3, \dots (X'\mathrm{X})^{-1}X'y_h)$$

$$= ((X'\mathrm{X})^{-1}X'lny_1, (X'\mathrm{X})^{-1}X'lny_2, (X'\mathrm{X})^{-1}X'lny_3, \dots (X'\mathrm{X})^{-1}X'lny_h)$$

$$\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots \hat{\beta}_h)$$

3.2.5 Test for Subsets of Regressors

This section tests whether or not the variable added to the model contributes significantly to the variation in the response variable. In testing whether or not the parameter β_i = 0 for each linear regression model in the stepwise process, is expressed as:

Null Hypothesis (H_0): $\boldsymbol{\beta_i}$ = 0, versus Alternative Hypothesis (H_1): $\boldsymbol{\beta_i} \neq 0$.

The test statistics is given as:

$$t_i^* = \frac{\beta_i}{s(\beta_i)}$$

Where $s(\beta_i)$ = the standard error of the estimate of β_i.

The estimator for the standard error of the estimate is $s(\beta_i) = \sqrt{\frac{SSE}{n-k-1}}$

Where SSE = SST – SSR and SST = the total sum of squares (Milton and Arnold, 1986).

The stepwise process reject the null hypothesis (H_0) if $t^* > t_{\alpha, (k, n-k)}$ or if p-value = P $(t_{\alpha, (k, n-k)}) < \alpha$, otherwise it fails to reject.

3.2.6 Test of Overall Regression (Lack-of-fit)

The overall regression hypothesis that none of the regressors (that is, the household characteristics in the model) predict the response variable (that is, household expenditure) is expressed as:

Null Hypothesis (H_0): $\boldsymbol{\beta} = (\beta_1, \beta_2, \ . \ . \ . \ , \beta_k) = 0$, versus

Alternative Hypothesis (H_1): At least one of the regressors predicts the response variable.

The test does not included $\beta_0 = 0$, so as not to restrict the model to have an intercept of zero (Alvin C. Rencher, 2002). The test statistic is defined as:

$$F^* = \frac{SSR/k}{SSE/n-k-1} = \frac{MSR}{MSE}$$

Which is distributed as $F_{k,\ n\text{-}k\text{-}1}$ when the null hypothesis (H_0: $\boldsymbol{\beta} = 0$) is true.

SSR = Regression Sum of Squares, SSE = Error Sum of Squares, MSR = Mean Square Regression, MSE = Means Square Error, k = the number of parameters to be estimated and n = the number of observations.

We reject the null hypothesis (H_0) if $F^* > F_{\alpha,\ (k,\ n\text{-}k\text{-}1)}$ or if the p-value = P $(F_{\alpha,\ (k,\ n\text{-}k\text{-}1)}) < \alpha$ otherwise we fail to reject it (Kutner et al, 2005).

3.2.7 Test of Model Assumptions

According to Kutner et al, 2005, model validity refers to the stability and reasonableness of the regression coefficients, the plausibility and usability of the regression function, and the ability to generalize inferences drawn from the regression analysis. The most preferred method to validate a regression model is through collection of new data sets. However, this is neither practicable nor feasible in this study, since funding is not sought for this academic work. Validation is a useful and necessary part of the model-building process. According to Torres-Reyna, 2007, how good the model is will depend on how well it predicts Y, the linearity of the model and the behavior of the residuals.

The tests of model assumptions exclude testing for multicollinearity since a stepwise regression technique used in developing the model controls for multicollinearity by ensuring that correlation between pairs of the predictor variables are not significant. The test also excludes testing for outliers because sample weight is applied to the data, and also the summary statistics shown in Table 4.8 in Appendix A, clearly indicates that there are no outliers in the data set.

The study performs the following diagnostics to determine whether or not the model departs from the model assumptions:

i. Test for normality: Residual plot is used to determine the normality of the error terms. The model assumes that the error term is normally distributed with mean 0 and constant variance.

   Let $\boldsymbol{e_i}$ represent the residual which is defined as $\boldsymbol{e_i} = Y_i - \hat{\beta}_i X_{ij} = Y_i - \hat{Y}_i$ where Y_i is the observed values of the regression and $\hat{Y}_i = \hat{\beta}_i X_{ij}$ is the predicted values of the regression.

   A histogram of the residuals indicates whether or not the error terms are normally distributed. If the error term is normally distributed then the histogram becomes symmetrical (bell-shaped), otherwise it becomes skewed. Residual values are defined as (m_error) in the analysis.

ii. Test for independence of error terms: Durbin-Watson test for independence of the error term is used. The Durbin-Watson test requires that the error terms are normally distributed with mean 0 and constant variance. Also, the test can be applied only when the regressors are strictly exogenous. A regressor X is strictly exogenous if Corr (X_s; E_t) = 0 for all s and t, which precludes the use of the Durbin–Watson statistic with models where lagged values of the dependent variable are included as regressors. (Kutner et al, 2005).

Let H_0 be the null hypothesis and H_1 be the alternative hypothesis. Also let ρ be the autocorrelation parameter, e_i is the residuals or the errors and n is the number of observations. The study tests the hypothesis

H_0: $\rho = 0$ meaning no first order autocorrelation, against H_1: $\rho > 0$ meaning positive first order autocorrelation.

The test Statistic is defined as: $$D = \sum_{i=2}^{n} \frac{(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e^2}$$

The decision rule is:

Rule 1: If $D > d_u$, then do not reject H_0. Where d_u is the upper bound at $\alpha = 5\%$ significance level.

Rule 2: If $D < d_l$, then reject H_0. Where d_l is the lower bound at $\alpha = 5\%$ significance level.

Rule 3: If $d_l < D < d_u$ then leads to no conclusion about whether or not to reject H_0.

iii. Random-Versus-Fitted plot (rvfplot) is applied to test for linearity of the regression function. The residuals values are plotted against fitted values in the Random-Versus-Fitted plot. Also a scatter plot of observed responses (lnrpcexp) and predicted responses (yhat) is used to confirm linearity of the regression function. If the plots indicate a linear pattern then the estimated regression model is linear. Any other pattern requires that the model be transformed or suggests that a nonlinear approach is desirable.

iv. Test for heteroskedasticity: "The error term $\boldsymbol{e_i}$ is homoskedastic if the variance of the conditional distribution of $\boldsymbol{e_i}$ given Xi (i.e. var $(\boldsymbol{e_i}|X_i)$), is constant for *i*=1…n, and, in particular, does not depend on X; otherwise, the error term is heteroskedastic" (Stock and Watson, 2003). Breusch-Pagan test is employed to test for homoscedasticity of the error terms. The Breusch-Pagan/Cook-Weisberg test for heteroskedasticity assumes that the error term of the estimated regression is normally distributed. When this condition is satisfied, we test the null hypothesis Null hypothesis (Ho): Constant variance against alternative hypothesis (H_1): Non-constant variance

The test statistics is given by $\chi^2_{bp} = \frac{\frac{SSR*}{q}}{\left(\frac{SSE}{n}\right)^2}$ where SSR* is the Sum of Square Regression of e_i^2 on X_i , SSE is the Sum of Square of the error term, n is the number of observations and q is the parameters. The test statistic (χ^2_{bp}) is approximately chi-square with q degree of freedom. The Decision rule is that if $\chi^2_{bp} > \chi^2_q$ or if p-value < 0.05 then the null hypothesis is rejected, otherwise we fail to reject the null hypothesis.

v. The lack of fit test is carried out to check how well the model fits the data. Lack of fit test is used to check if the linear regression model is appropriate for the data. This test procedure is only possible if repeated responses are available at some values of X, and the responses Y for some given X are independently normally distributed with constant variance σ^2. The lack-of-fit test is the test for the overall regression. The details involved have already been mentioned above (Test of overall regression).

Test for Omitted-Variable Bias: Omitted-variable bias test is performed to check whether important variables are not present in the model.

In this case we test the Null hypothesis (H_o): model has no important variable omitted against Alternative hypothesis (H_1): model has omitted important variables.

According to Torres-Reyna, 2007, testing for omitted variable bias is important for modeling since it is related to the assumption that the error term and the independent variables in the model are not correlated ($E(e|X) = 0$).

The omitted-variable bias test fits the regression $Y_i = X_i b + Z_i t + U_i$ and then performs a standard F test of $t = 0$. The default test uses $Z_i = (b\hat{y}_i^2, b\hat{y}_i^3, b\hat{y}_i^4)$. If $(b\hat{y}_i^2, b\hat{y}_i^3, b\hat{y}_i^4)$ is specified, then $Z_i = (x^2_{1i}; x^3_{1i}; x^4_{1i}; x^2_{2i}; \ldots.. x^4_{mi})$. In either case, the variables are normalized to have minimum 0 and maximum 1 before the powers are calculated. If the test shows a p-value less than the significance level α or if $F^* > F(\alpha, n\text{-}k, n\text{-}1)$ then the null hypothesis is rejected. (Goldstein 1991, 1992)

3.2.8 Remedial Measures

If any of these tests reject the null hypothesis, then inferences from the model would not be reliable. Therefore, in such a situation, remedial measures such as Box-Cox transformation is applied to ensure that the model do not depart from its assumptions. The transformation ensures normality, and the final model is also weighted to take care of nonlinearity and influential outliers. Alternatively, a Generalized Least Squares Model is used when the error term in not independent. The stepwise regression technique eliminates the effects of multicollinearity.

3.3 MODEL VALIDATION

According to Kutner et al, 2005, model validity refers to the stability and reasonableness of the regression coefficients, the plausibility and usability of the regression function, and the ability to generalize inferences drawn from the regression analysis. The most preferred method to validate a regression model is through collection of new data set. However, this is neither practicable nor feasible in this study.

Validation is a useful and necessary part of the model-building process. There are several methods of assessing model validity, but this research employs the K-fold cross-validation method of model assessment.

With this method, the GLSS6 data set is first split into 10 parts of equal or similar characteristics. For k = 1, 2, ..., 10, one part is used as the validation set and, the other nine parts are used to fit the model. The nine parts used for building the model is called the model sample (or training data) and the other part used to test out of sample model performance is referred to as validating sample (or testing data). The process is simulated 20 times for each fold until all the ten folds are exhausted. After the validity of the model is assessed, the final model is built on the whole data set. The diagram below illustrates random splitting of the GLSS6 data set into ten-folds and the process of cross-validation.

Step 1: Randomly split the full data set into ten folds having similar characteristics.

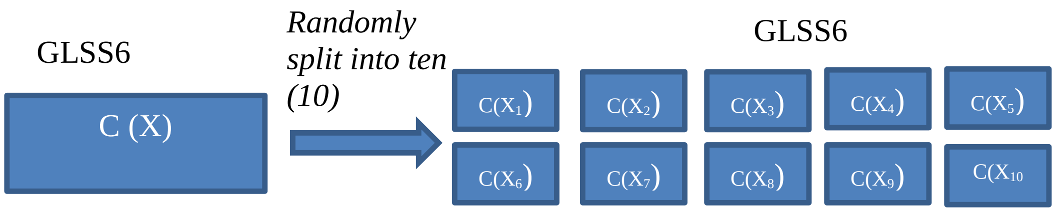

Figure 3.1: Full GLSS6 data set randomly split into ten similar sub-data sets

Step 2: Create nine possible training data for modeling and a testing data for each of the 10 randomly split data.

Training Data $C\ (X_i)$ Testing Data $C\ (X_j)$

Where $i = 1, 2, \ldots, 10$ and $j = 1, 2, \ldots, 10$ for $i \neq j$. For example when $j = 1$, $i = 2, \ldots, 10$. When $j = 2$, $i = 1, 3, \ldots 10$ and when $j = 3$, $i = 1, 2, 4, \ldots, 10$ and when $j = 10$, $\mathrm{i} = 1, 2, \ldots, 9$.

Step 3: Create the model by stepwise regression technique with nine folds (training data) and test it for out of sample performance with the other data set (testing data) remaining. This is repeated ten times and the average of all estimates obtained.

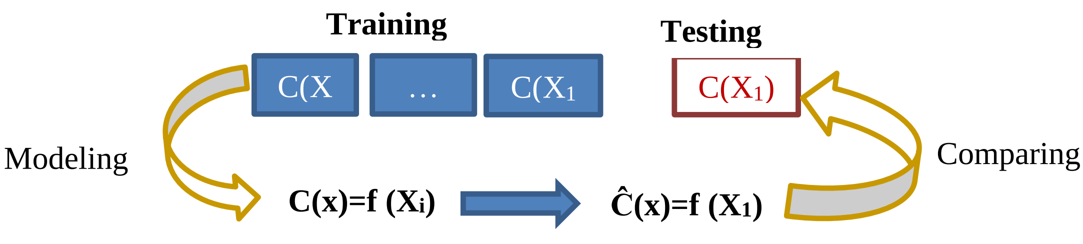

Figure 3.2: Modeling with training data and testing with testing data.

The 10 estimates of prediction error are then combined to produce a 10-fold cross-validation estimate. The statistics of interest are the Mean Square Error (MSE), Average Predicted Poverty Head Count and the Absolute (or Squared) Differences (ASD) between the actual and projected poverty rates.

These statistics are mathematically written as:

MSE = $\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2$ ASD = $\frac{1}{10}[\sum_{i=1}^{10}(H_i - \hat{H}_i)^2]$, where:

Y_i = the value of the response variable in the ith validation case,

$\hat{Y}_i$ = the predicted value for the i^{th} validation case based on the model-building data set,

N = the number of cases in the validation data set,

H_i = Actual poverty rate in the i^{th} response validation case

$\hat{H}_i$ = Predicted poverty rate for the i^{th} validation case based on the model-building data set

3.4 MULTIPLE IMPUTATION

Multiple Imputation (MI) is a statistical technique for analyzing incomplete data sets, that is, data sets for which some entries are missing. It provides a useful strategy for dealing with data sets with missing values. Instead of filling in a single value for each missing value, Rubin's (1987) multiple imputation procedure replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute (Yuan, 2011). The application of this technique requires three steps: Imputation, Analysis and Pooling. Before the three steps, the data set is examined to determine the pattern of missingness. The missing data pattern can be monotonic or non-monotonic. Based on the missing data pattern, the imputation step is determined.

Step1: Imputation - Imputes (fills in) the missing entries of the incomplete data sets, not once, but m times (*m*=20 in this study). In other words, this stage imputes multiple values for each missing value in the data set to create *m* complete data sets. Imputation randomly selects the value to fill the missing value (that is, household expenditure in MICS4) from the predictive distribution of missing data given observed data (i.e. household expenditure in GLSS6).

Step 2: Analysis - Analyses each of the *m* completed data sets. This step performs *m* separate analyses. That is, analyzing each completed data set separately to generate *m* sets of estimates. Every set of estimates may differ slightly from each other.

Step 3: Pooling - Integrates the *m* analyses results into a final result. That is, combining all sets of estimates to generate an overall estimate and calculating the variation among parameter estimates.

Rubin (1987) shows that if the method used to create the imputation is suitable, then the resulting inferences will be statistically valid. The most challenging step is Imputation, that is, the construction of the *m* completed data sets. This step accounts for the process that created the missing data. Typical problems associated with imputations are: The fact that something is missing could be related to its value (e.g., people with higher incomes tend to skip income questions more often); missing entries can appear anywhere in the data; the method used for the imputations must be related to the intended complete-data analyses. The repeated analysis step on the imputed data is actually somewhat simpler than the same analysis without imputation, since there is no need to bother with the missing data. The pooling step consists of computing the mean over the *m* repeated analysis, its variance, and its confidence interval or p-value. In general, these computations are relatively simple (Van Buuren, 2012).

There are several methods of multiple imputations (Yuan, 2011). These include the Markov Chain Monte Carlo (MCMC) method (Schafer, 1997), Propensity Score (PS) method (Rosenbaum and Rubin, 1983) and the Regression Multiple Imputation method (Rubin, 1987). For a data set with a monotone missing data pattern, one of the following methods can be used: a regression method, a predictive mean matching method (Heitjan and Little, 1991; Schenker and Taylor, 1996), or a propensity score method (Lavori, Dawson, and Shera, 1995) to impute missing values for a continuous variable (Yuan, 2011). This study uses the regression multiple imputation method by Rubin to impute missing values (that is, household expenditure for MICS4 data set), which is a continuous variable.

With this method, the response variable (MICS4 households' expenditure) of the final model consists of only missing values. At the initial stage, a model is fitted using observations with observed values (that is, expenditure values of the GLSS6 households) for the response variable. With this observed model from the GLSS6, a new model is drawn, and is used to impute missing values sequentially for the response variable (household expenditure) in another data set (MICS4). That is, for a variable Y_j (household expenditure) in another data set (MICS4) with all the values missing. Therefore, the missing values are imputed from the distribution $Yj \sim P\ (Y_j \mid Y_1;\ Y_2,\ \ldots\ldots,\ Y_{j-1})$ of the observed values (i.e. household expenditure and the covariates from the GLSS6 data set). According to Yuan (2011), given a model in the form:

$Yj = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k$ where $X_1, X_2, \ldots, X_k$ are the covariates generated from preceding variables $Y_1; Y_2, \ldots, Yj_{-1}$, the following steps are used to impute missing values for Y_j at each imputation:

1. The regression model is fitted using observed values for the variable Y_j and its covariates $X_1, X_2, \ldots, X_k$. The fitted model includes the regression parameter estimates $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ and the associated covariance $\hat{\sigma}_j^2 V_j$ where $V_j = (X'X)^{-1}$ is the inverse matrix derived from the intercept β_0 and covariates $X_1, X_2, \ldots, X_k$.

2. New parameters $\beta_* = (\beta_{*0}, \beta_{*1}, \ldots, \beta_{*k})$ and σ_{*j}^2 are drawn from the posterior predictive distribution of the parameters (Rubin, 1987). That is, they are simulated from $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$, $\hat{\sigma}_j^2$ and $(X'X)^{-1}$. The variance is calculated as $\sigma_{*j}^2 = \hat{\sigma}_j^2 (n_j - k - 1)/g$ where $g = \chi^2_{n_j - k - 1}$, is a random variate and n_j is the number of non-missing observations for Y_j. The regression coefficients are also calculated as $\beta_* = \hat{\beta} + \hat{\sigma}_j^2 V'_{hj} Z$ where V'_{hj} is the upper triangular matrix in the Cholesky decomposition, $V_j = V'_{hj} V_{hj}$ and Z is a vector of $k + 1$ independent random normal variate (Yuan, 2011).

3. The missing values are then replaced by $b_{*0} + b_{*1}x_1 + b_{*2}x_2 + \cdots + b_{*k}x_k + z_j \sigma_{*j}$ where $x_1, x_2, \ldots, x_k$ are the values of the covariates and z_j are simulated normal deviate.

Yuan (2011), stressed that the predictive mean matching method can also be used for imputation. The author states that it is similar to the regression method except that for each missing value, it imputes an observed value which is closest to the predicted value from the simulated regression model. The writer points out that the predictive mean matching method ensures that imputed values are plausible and may be more appropriate than the regression method if the normality assumption is violated as observed by (Horton and Lipsitz, 2001).

Combining Inferences from Imputed Data Sets: Rubin's Rules

With *m* imputations, we compute *m* different sets of the point and variance estimates for a parameter Q, where $\hat{Q}_i$ and $\hat{U}_i$ are the point and variance estimates from the *i*th imputed data set, *i*=1, 2, ..., m. Then the point estimate for Q from multiple imputations is the average of the *m* complete-data estimates given as:

$$\bar{Q} = \frac{1}{m}\sum_{i=1}^{m}\hat{Q}_i$$

Let $\bar{U}$ be the within-imputation variance, which is the average of the *m* complete-data estimates and B be the between-imputation variance. Therefore,

$$\bar{U} = \frac{1}{m}\sum_{i=1}^{m}\hat{U}_i$$

and $$B = \frac{1}{m-1}\sum_{i=1}^{m}(\hat{Q}_i - \bar{Q})\,(\hat{Q}_i - \bar{Q})^T$$

Thus, the variance estimate associated with Q is the total variance given as

$$T = \bar{U} + (1 + \frac{1}{m})B$$

The statistic $(Q - \bar{Q})T^{-1/2}$ is approximately distributed as a t-distribution with *Vm* degrees of freedom (Rubin 1987), where

$$V_m = (m-1)\left[1 + \frac{\bar{U}}{(1+m^{-1})B}\right]^2$$

When the complete-data degrees of freedom V_0 is small and there is only a modest proportion of missing data, the computed degrees of freedom, V_m, can be much larger than V_0, which is inappropriate. Barnard and Rubin (1999) recommend the use of adjusted degrees of freedom, V_m^*,

$$v_m^* = \left[\frac{1}{v_m} + \frac{1}{\hat{v}_{obs}}\right]^{-1}$$

$$\text{Where} \quad \hat{v}_{obs} = \frac{v_0+1}{v_0+3}\, v_0\left(1 - \frac{(1+m^{-1})}{T}B\right)$$

3.5 DATA MANAGEMENT AND ANALYSIS

The GLSS6 and MICS4 data sets obtained from the Ghana Statistical Service were prepared in STATA format. The study used version 12 of the STATA software for data management and analysis. Most of the graphs are generated using the 2010 version of Microsoft Excel. The other graphs are obtained from the STATA analysis system.

Data management included recoding response option of identical variables found in the two data sets to match. Some identical variables found were renamed to give the same name to them, and other variables were derived, where possible, from two or more other variables to match a corresponding variable in the other data set. Dummy variables were generated from the categorical variables in each data set for regression purposes. Frequency tables were generated to examine the number missing observations in each variable. Since using multiple imputation technique to impute missing values is not the objective of this study, variables with too many missing values were not included in the analysis.

The analysis was done in three parts. The first part was the preliminary analysis which focuses on descriptive statistics. These are cross-tabulations of household characteristics and poverty status to identify how these characteristics are associated with poverty. The second part applies stepwise regression technique to select variables and estimate parameters for the model. The final part employs the application of multiple imputation technique to impute household expenditure values into the MICS4 data to determine incidence of poverty from the MICS4 data set using multiple imputation regression.

Prior to the application of the model and multiple imputation, a cross-validation of the model is performed to validate the performance of the model. The estimates from the cross-validation results were compared with those of the GLSS6 to validate the model. After application of the model to MICS4 data, the incidence of poverty estimated was compared with the GLSS6 estimate to ascertain closeness of the estimates.

3.6 VARIABLE DESCRIPTION

Table 3.3 describes the household expenditure (rpcexp), and the other household variables that are common to both GLSS6 and MICS4 data sets used for modeling.

Table 3.3: List of household variables common to both surveys

Variable Name	type	Format	Variable Label
rpcexp	float	%9.0g	Household expenditure
urban	float	%9.0g	Locality of residence: Urban = 1, Rural = 0
region	byte	%13.0g	Region of residence
hhsize	byte	%8.0g	Household size
head_male	float	%9.0g	Household head is male = 1, 0 otherwise
head_age	float	%9.0g	Age of the household head
head_schooling	float	%27.0g	School attendance of the household head
schlvl	float	%26.0g	Highest level of schooling completed by the household head
read	float	%9.0g	Household head can read a simple sentence in any language
write	float	%9.0g	Household head can write a simple sentence in any language
rooms	float	%9.0g	Number of rooms
bedrooms	float	%9.0g	Number of bedrooms
tenure	byte	%30.0g	Tenancy arrangement
water_drinking	byte	%94.0g	Main source of drinking water
fuel	byte	%67.0g	Main source of cooking fuel
water_general	byte	%94.0g	Main source of water for general use
wall	byte	%75.0g	Main construction material of outer wall
toilet	byte	%32.0g	Type of toilet
floor	byte	%86.0g	Main construction material of floor
roof	byte	%63.0g	Main construction material of roof
fridge	float	%9.0g	Any household member or head owns a refrigerator
radio	float	%9.0g	Any household member or head owns a radio
desktop	float	%9.0g	Any household member or head owns a desktop computer
laptop	float	%9.0g	Any household member or head owns a laptop computer
tv	float	%9.0g	Any household member or head owns a television
bicycle	float	%9.0g	Any household member or head owns a bicycle
mcycle	float	%9.0g	Any household member or head owns a motor cycle
land	float	%9.0g	Any household member or head owns a land
car	float	%9.0g	Any household member or head owns a car
washm	float	%9.0g	Any household member or head owns a washing machine
cdplayer	float	%9.0g	Any household member or head owns a CD/DVD player

CHAPTER FOUR

ANALYSIS AND RESULTS

4.0 INTRODUCTION

This chapter presents the variables of interest, and how these variables are related to poverty in order to identify the candidate variables for the model. It also focuses on developing a model from the GLSS6 data set using stepwise regression method, and by regression multiple imputation method and the model developed from the GLSS6 data set, impute household expenditure and estimate poverty incidence from the MICS4 data set.

The predictor variables of interest are grouped under location, demography, education, dwelling characteristics and household assets ownership. The response variable is the log of annual household expenditure. In the MICS4 data set, the response variable is assumed to be missing for all observations. Multiple imputations is then used to estimate the expenditure values for the observations and incidence of poverty calculated at national as well as for urban and rural levels.

This model building process involves performing diagnostics tests (i.e. testing the assumptions underlying the model); estimating the final model; and finally, applying the model to impute household expenditure for the MICS4 data set.

The categorical variable among the predictors are recoded into binary variables for modeling purpose. These binary variables are presented in Table 5.1 in Appendix A.

Examples of the binary variables created for tenure, school attendance, main material for floor and main source of lighting are as follows.

<u>Tenure</u>

tenure1 = 1, Owning
0, Otherwise

tenure2 = 1, Renting
0, Otherwise

tenure3 = 1, Rent free, perching etc.
0, Otherwise

<u>Main material of floor</u>

floor1 = 1, Earth/mud
0, Otherwise

floor2 = 1, Cement/concrete
0, Otherwise

floor3 = 1, Wood, tiles, terrazzo
0, Otherwise

<u>School Attendance of household head</u>

head_schooling1 = 1, Never attended
0, Otherwise

head_schooling2 = 1, Still attending school
0, Otherwise

head_schooling3 = 1, Attended school in the past
0, Otherwise

<u>Main Source of lighting</u>

lighting1 = 1, Electricity or Solar
0, Otherwise

lighting2 = 1, Kerosene or gas lamp
0, Otherwise

lighting3 = 1, Flashlight/torch
0, Otherwise

lighting4 = 1, Candle, firewood etc.
0, Otherwise

4.1 PRELIMINARY ANALYSIS

The preliminary analysis provides descriptive statistics on poverty and household characteristics. It comprises cross-tabulations and graphs of poverty status and other household characteristics, which are identified to be common to the GLSS6 and MICS4 data sets. Considering the six rounds of Ghana Living Standard Survey conducted by the Ghana Statistical Service, it is observed that each round of the survey calculates a unique national poverty line and estimates the population with annual total expenditure below that line to determine poverty incidence or headcount.

4.1.1 Trend in Poverty Incidence and Poverty Gap Ratio

i. Incidence of poverty

Incidence of poverty, which is also known as poverty headcount is a measure of the proportion of population that is poor (GSS, 2014). According to the GLSS6, 24.2 percent of the total population of Ghana is poor. In other words, 24.2 percent of Ghana's population has their total annual expenditure below the national poverty line of GH₵ 1,314.00 (One thousand, three hundred and fourteen Ghana Cedis) for 2012/13.

The recent rounds of GLSS, between 1991/92 (GLSS3) and 2012/13 (GLSS6) have shown a decreasing trend in the percentage of population living below the national poverty line as shown in Figure 4.1. The graph shows higher rate of reduction in poverty headcounts or incidences between 1991/92 and 1998/99 and between 1998/99 and 2005/06 than between 2005/06 and 2012/13. The figure shows almost the same rate of reduction in poverty incidence between 1991/92 and 2005/06, whereas the rate of reduction in poverty incidence between 2005/06 and 2012/13 slackened significantly.

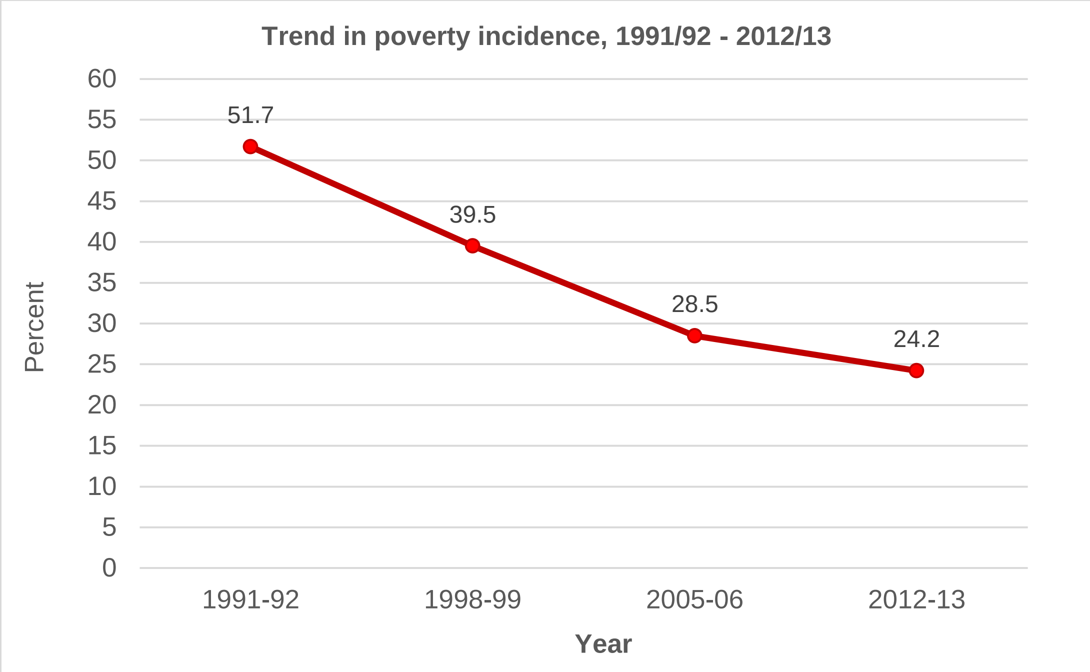

Figure 4.1: Trend in incidence of poverty in Ghana, 1991/92 – 2012/13

ii. Incidence of poverty by region

The regional distribution of poverty incidence also shows a generally decreasing trend, though some regions had poverty incidence fluctuating between the years. Figure 4.2 shows that the three upper regions have very high proportions of the population being poor in all of the years. The region with the lowest proportion of the population being poor is Greater Accra.

Though Greater Accra region has the largest population size, it has the lowest poverty incidence in each of the years. This means that Greater Accra is the region with the lowest proportion of its population being poor in spite of its being the region with the largest population in Ghana according to the 2010 Population and Housing Census.

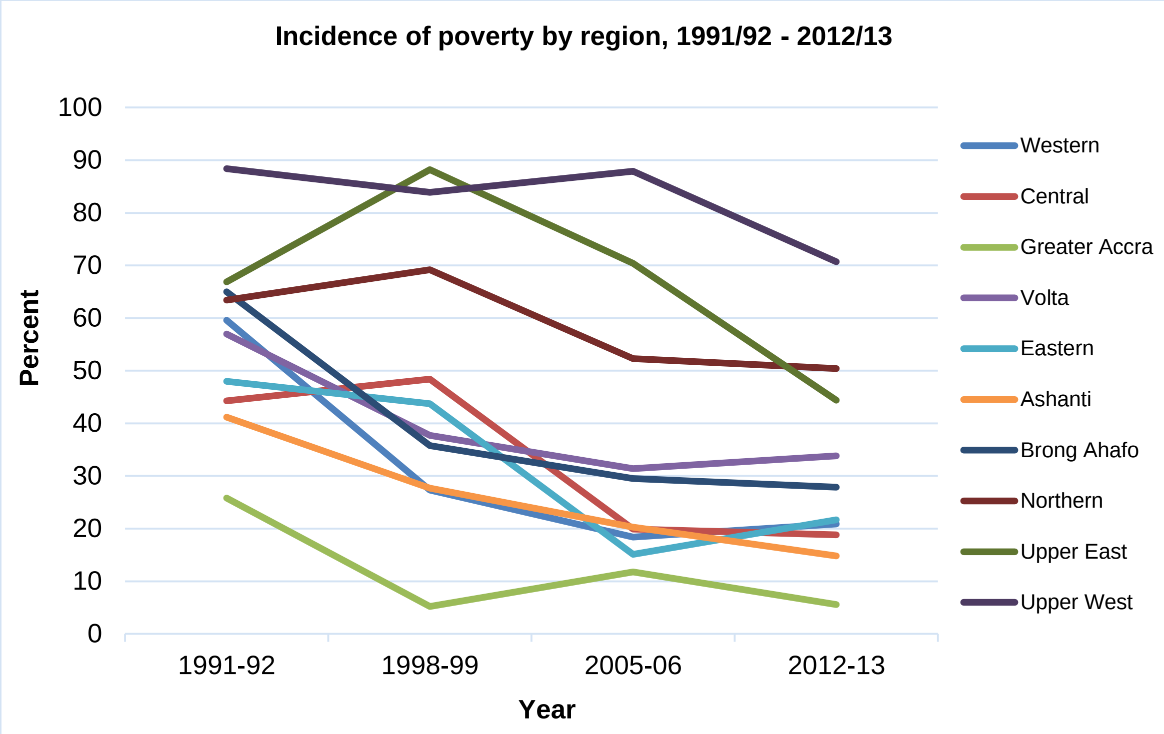

Figure 4.2: Trend of poverty incidence by region, 1991/92 – 2012/13

iii. Poverty gap ratio

Poverty gap ratio is the average ratio of the gap to which individuals fall below the national poverty line, hence it is zero for the non-poor. It measures the intensity of poverty in the country. This index takes into account both incidence and depth of poverty. It gives an indication of the minimum level of resources which is required to eliminate poverty, assuming that resources could be perfectly targeted to raise the poor exactly to the poverty line. (GSS, 2014)

The poverty gap ratio in Ghana follows similar trend as the poverty headcount. According the various rounds of GLSS conducted by the Ghana Statistical Service, poverty gap ratio has dropped considerably from 18.5 percent to 7 percent. Figure 4.3 depicts the trend of the poverty gap ratio from 1991/92 to 2012/13.

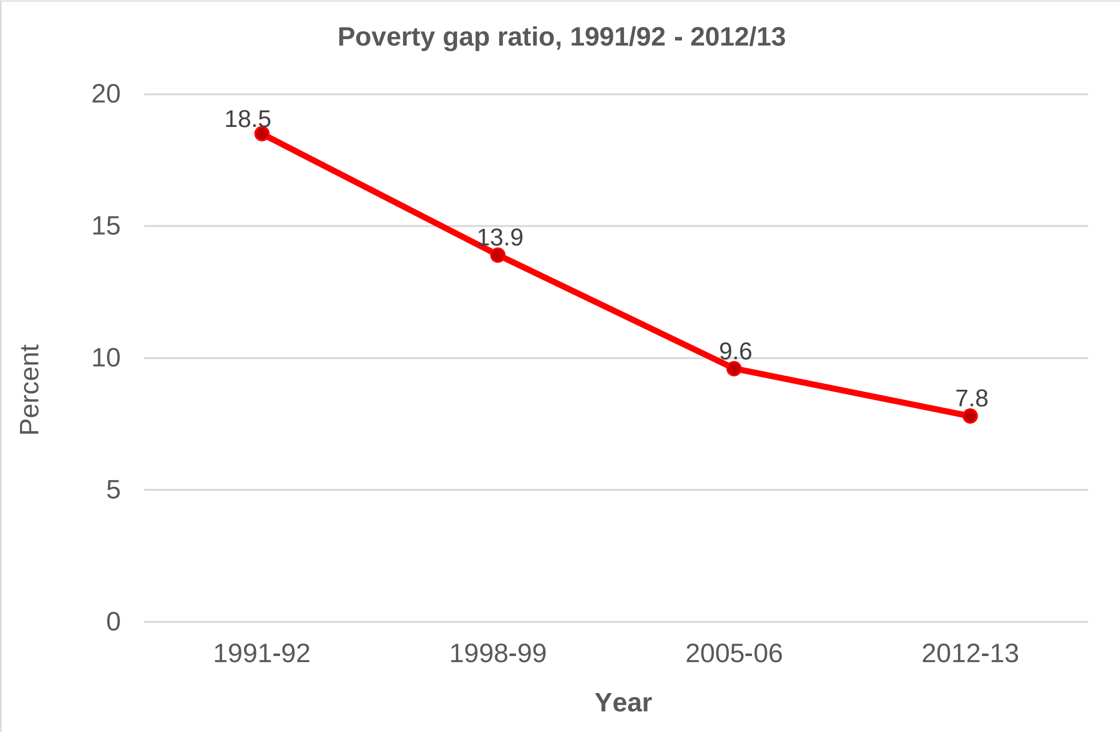

Figure 4.3: Trend of poverty gap ratio, 1991/92 – 2012/13.

iv. Poverty gap ratio by region

The regional distribution of poverty gap ratio also showed a general decrease over the period in question. The three upper regions (that is Northern, Upper East and Upper West) showed higher poverty gap ratios than the other seven regions in Ghana as shown in Figure 4.4.

Greater Accra, like the poverty incidence, recorded the lowest poverty gap ratio over the entire period.

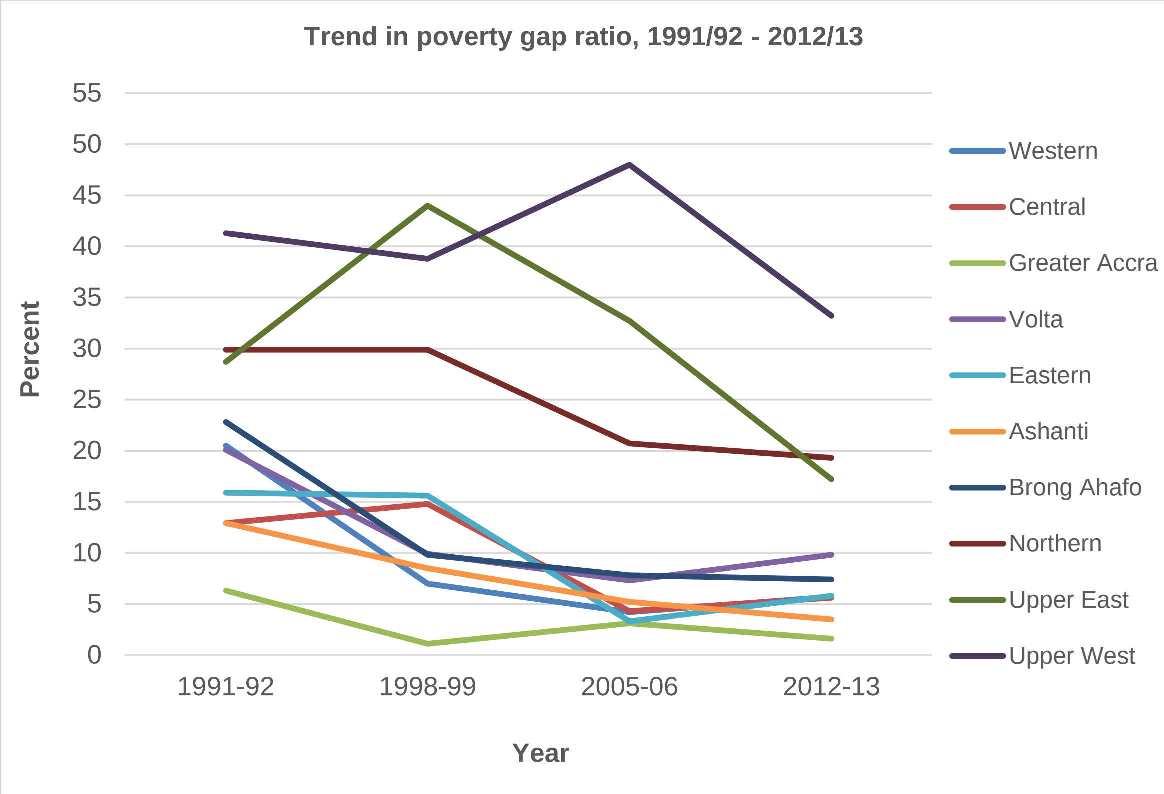

Figure 4.4: Trend in poverty gap ratio by region, 1991/92 – 2012/13

Though Greater Accra region recorded lowest proportion of population living below the national poverty line and the lowest poverty gap ratio, it is the region with the highest population. This means that in terms of numbers, it still has greater number of persons who are poor. On the other hand, the population of each of the three upper regions is evidently less than the other seven regions in spite of the fact that they have higher proportion of poor persons. Therefore, in terms of numbers, these three regions have lesser poor persons in the country.

v. Incidence of poverty by locality of residence

Locality of residence has significant effect on the standard of living of the residence as shown by Figure 4.5 and Table 4.1 (Page 73). Though the graph shows a decreasing trend in incidence of poverty, the figures recorded for households in urban localities are far lower than those in rural localities. Unlike between 1991/92 and 2005/06, there is a small percentage difference in the poverty incidence recorded between 2005/06 and 2012/13 for both urban and rural localities.

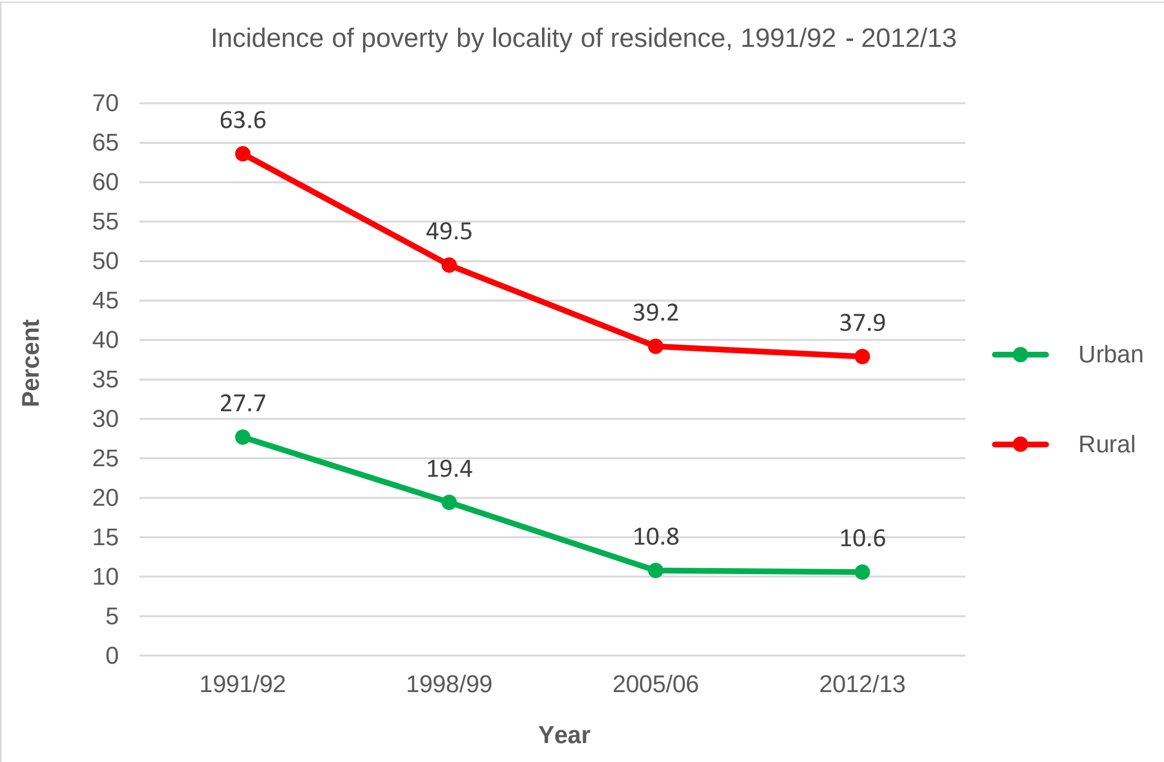

Figure 4.5: Trend of poverty incidence by locality of residence, 1991/92 – 2012/13

Table 4.1: Proportion of household characteristics by poverty status and household

	Poverty Status		
	Poor	Non-poor	All
Locality of residence			
Urban	21.9	59.1	50.1
Rural	78.1	40.9	49.9
Total	100	100	100
Age of household head (1)			
Working age (15-64 years)	20.2	67	87.2
Non-working age (65+ years)	4.1	8.8	12.8
Total	24.2	75.8	100
Age of household head (2)			
Working age (15-64 years)	83.3	88.4	87.2
Non-working age (65+ years)	16.7	11.6	12.8
Total	100	100	100
Household Size			
Small household (1-4 members)	16.6	44.2	37.5
Large household (5+ members)	83.4	55.8	62.5
Total	100	100	100
Main toilet facility			
No facility	45.7	15.5	22.8
Improved	6.1	28.8	23.3
Unimproved	48.2	55.7	53.9
Total	100	100	100
Occupancy Condition			
Overcrowded	69.3	54.1	57.8
Spacious	30.7	45.9	42.2
Total	100	100	100
Asset Ownership			
Own no asset	15.2	7.2	9.1
Own asset(s)	84.8	92.8	90.9
Total	100	100	100

Source: GLSS6, Ghana Statistical Service

Table 4.1 shows that higher proportion of the non-poor households (59.1%) reside in urban localities whereas higher proportions (78.1%) of the poor households reside in rural localities. In other words, almost two-fifth of the non-poor household live in rural localities while almost a fifth of the poor household heads live in urban localities.

4.1.2 Household characteristics and poverty status

This section considers certain household characteristics against poverty status using the GLSS6 data set. These characteristics are common to both GLSS6 and MICS4 data sets. They include age of head of household, highest educational level of household head, household size, main source of drinking water, type of toilet facility used and main source of fuel for cooking.

These household characteristics are empirically known to have significant relationship with poverty irrespective of the method of measurement. Tables on the relationship between these characteristics are presented in Appendix A. These tables have the same number as the figure but suffixed with the letter “a”.

i. Age of household head by poverty status

It is observed from the data set that the minimum age of a household head is 15 years and the maximum age is 99. In the interviewers manual used during data collection, age 99 is a code which represent age 99 years and older. Table 4.1 indicates that majority of household heads, 87.2 percent, fall within the working age (15-64 years) group. The remaining 12.8 percent are 65 years or older. It is also observed that 24.2 percent of household heads are poor whereas 75.8 percent are non-poor.

From Table 4.1, 83.3 percent of the poor household heads fall within the working age group whereas the other 16.7 percent are 65 years or older. Among those that are non-poor, 88.4 percent are in the working age category and 11.6 percent are 65 years or older.

ii. Household size by poverty status

The analysis divides household size into two groups, small and large households. The small households fall below the national average household size of 4.5 according to the 2010 PHC, and the large households have members above the national average. Table 4.1 below indicates that 37.5 percent of households have household sizes between 1 and 4 while 62.5 percent have household size of 5 or more. It can be observed that a considerable proportion (44.2%) of the non-poor household have small household size, but among the poor households, 83.4 percent have large household sizes.

iii. Highest educational level of household head by poverty status

The analysis of the data set, as shown in Figure 4.6, reveals that the higher the level of education, the higher the proportion of household head who are non-poor. The opposite is also true. This indicates an inverse relationship between the level of education attained by household heads and their standard of living.

Figure 4.6 also indicates that almost all household heads with SHS or higher level of education are non-poor. It is further observed that greater proportion of household heads who are poor have attained up to primary level of education. However, an appreciable proportion of non-poor household heads have no education.

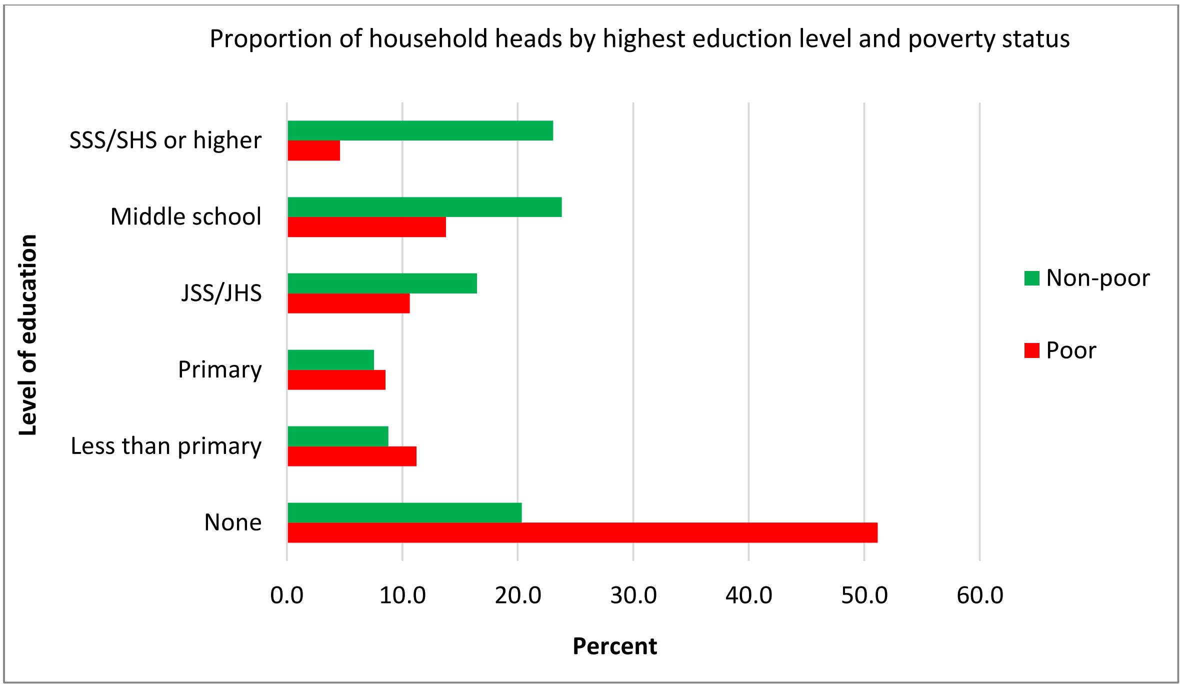

Figure 4.6: Proportion of household head by highest education level and poverty status

iv. Main source of drinking water by poverty status

The different sources of water listed in the questionnaire are grouped into safe and unsafe source of drinking water by exposure to external contamination. Piped water, whether inside or outside the dwelling, water from public stand pipe, borehole, tube-well, protected well, satchet and bottled water are classified as safe source of drinking water. The other sources which include rain water, water from tanker supply, unprotected well, lake, pond, and unprotected spring are classified as unsafe source of drinking water.

It is observed from Figure 4.7 that high proportion of non-poor households have access to safer source of drinking water. However, it can also be identified that there is a substantial number of non-poor households using unsafe source of drinking water.

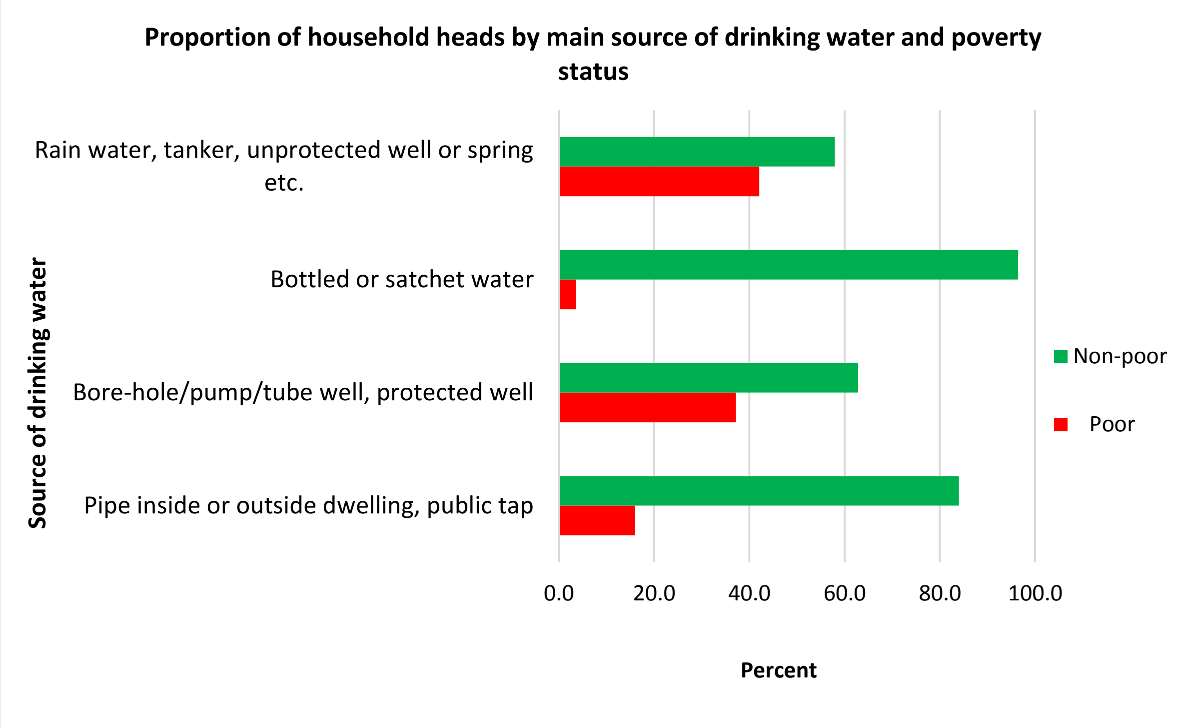

Figure 4.7: Proportion of household by source of drinking water and poverty status

v. Main source of cooking fuel by poverty status

The ten response options of the question, on main source of cooking fuel, are grouped into four categories. The grouping was based on frequency of each option. The four categories are wood, charcoal, gas and others (i.e. Electricity, kerosene, saw dust, animal waste, crop residue and other). Households that do not cook at all are drop out of the analysis.

Figure 4.8 reveals that a very small proportion of poor households use gas for cooking. Most of them use wood and/or charcoal for cooking. The non-poor households use gas, wood and charcoal for cooking. Only few use other sources of fuel such as electricity for cooking. Also, among the households using wood for cooking, the proportion of the poor households is far more than the non-poor household.

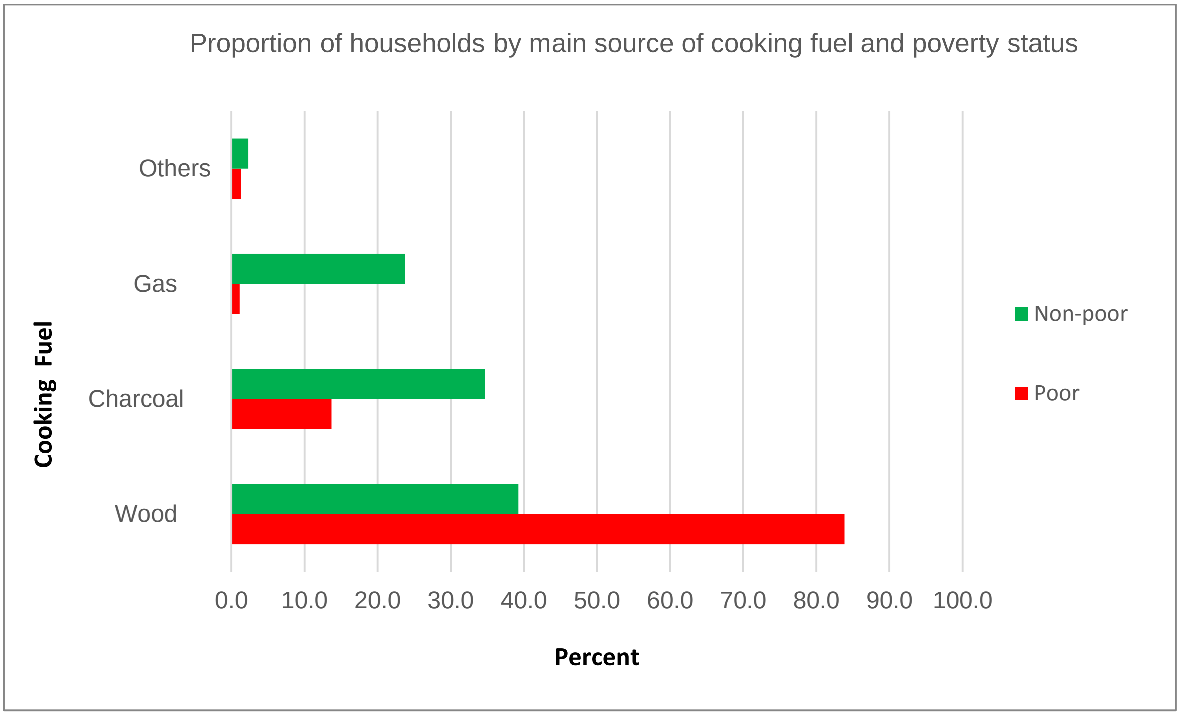

Figure 4.8: Proportion of households by main cooking fuel and poverty status

vi. Main toilet facility used by poverty status

The list of toilet facilities is grouped into no facility, improved and unimproved toilet facilities. The improved facilities are Water Closet (WC) and KVIP. The unimproved facilities are bucket/pan latrine, pit latrine (with slabs), public toilet and others.

Table 4.1 indicates that 28.8 percent of the non-poor households use improved toilet facilities whereas only 6.2 percent of the poor households use improved toilet facility. Among the households with no toilet facility, the poor households (45.7%) are about three times more than the non-poor households (15.5%). Greater proportion of both the poor (48.2%) and non-poor households (55.7%) use unimproved toilet facility.

4.1.3 Dwelling Characteristics and Poverty Status

The dwelling characteristics selected are common to both data sets. They include main material for construction of external wall, main floor material, main construction material of roof, number of sleeping rooms and tenure. These characteristics are crossed with poverty status to identify their relationships with poor and non-poor households.

i. Proportion of households by main external wall material by poverty status

The main construction material is first grouped into conventional and non-conventional wall materials. The conventional materials include stone, burnt bricks, cement blocks and concrete. The non-conventional materials are further divided into permanent and temporal materials. The permanent materials are wood, metal sheet/slate/asbestos and bamboo while the permanent materials are mud bricks and landcrete.

Figure 4.9 indicates that greater proportion of non-poor households use conventional material for construction of external wall whereas most of the poor households use non-conventional materials for construction of external wall.

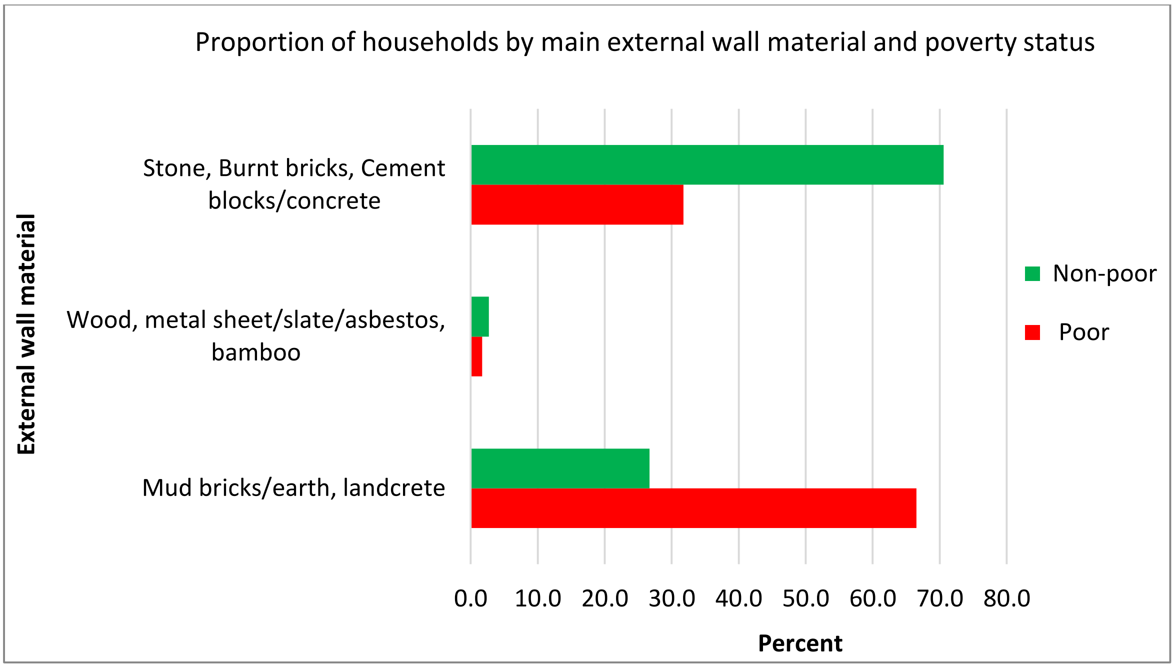

Figure 4.9: Proportion of households by main external wall material and poverty status

ii. Proportion of households by main material of floor by poverty status

The floor material is also grouped into three categories based on type and quality. Group one constitutes earth/mud: the lowest quality. Group two includes cement/concrete/ stone/burnt bricks: the next higher quality, and Group three is made up of polished wood /vinyl/ceramic/porcelain/granite/terrazzo/marble tiles: the best quality. The outcome of cross-tabulating these floor materials with poverty status is depicted in Figure 4.10.

It is observed from Figure 4.10 that most of the poor and non-poor households use cemented floor. About 88 percent of the non-poor households and 81 percent of poor households have floors made of cement. A very small proportion of non-poor households use tiled floor and about the same proportion of the non-poor households use earth or mud floor.

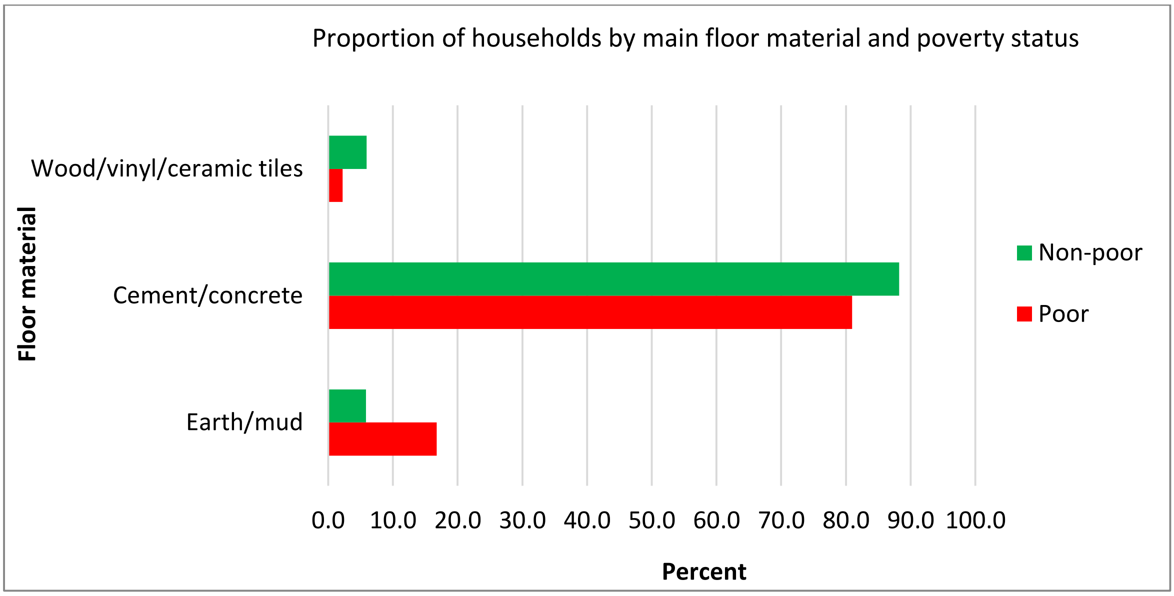

Figure 4.10: Proportion of households by main floor material and poverty status

iii. Proportion households by main material of roof and poverty status

The nine different materials for roofing are grouped into four categories in terms of quality for the analysis. These groups are displayed in the results shown in Figure 4.11. The figure clearly indicates that most of the poor (70.5%) and non-poor (78.2%) households use metal sheet for roofing.

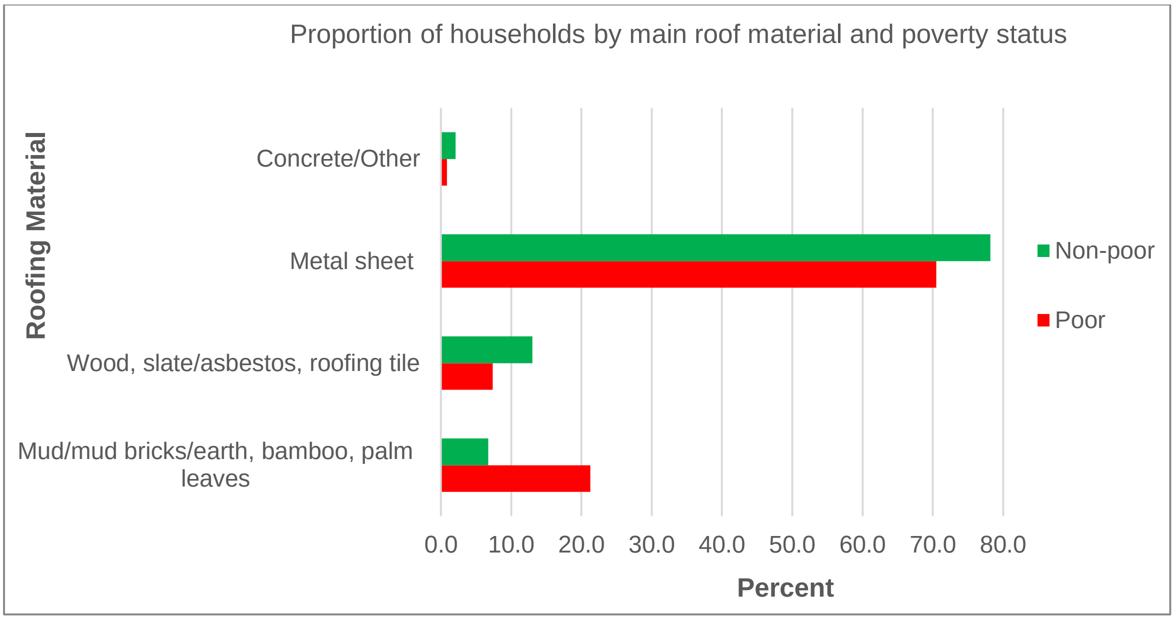

Figure 4.11: Proportion of households by main roofing material and poverty status

iv. Number of sleeping rooms by poverty status

It is empirically proven that overcrowding is directly related to poverty. The number of sleeping rooms per household helps measure congestion or overcrowding. According to the U.S. Department of Housing and Urban Development, 2007, overcrowding is commonly defined as is persons-per-room in a dwelling unit. The analysis adopts another definition of overcrowding as the total number of persons in a unit, regardless of unit size. In this analysis, if persons-per-room is less than 0.5 then the household is overcrowded. This means that at most 2 persons are expected to sleep in a room, and if persons sleeping in a room are more than 2 then there is overcrowding.

The result shown in Table 4.1 indicates that higher proportions of both the poor and the non-poor households are overcrowded. This means that there are more than two persons that sleep in one room in these households. The proportion of poor households suffering overcrowding conditions is higher than that of the non-poor households. However, it is also observed that a higher proportion of the non-poor households have spacious occupancy condition.

v. Proportion of households by tenure and poverty status

Tenure is the condition under which a dwelling unit is occupied. In other words it is tenancy or occupancy arrangement. Furthermore, it is the length of time that a household occupies a dwelling unit with or without agreed terms and payments of rent. Figure 4.12 displays the proportion of households by tenure and poverty status. The figure indicates that most of both poor households own the dwelling unit. About the same proportion of non-poor households in rent free dwelling units rent their dwelling units. There is also a significant proportion of poor households in rent free dwelling units. Only a small proportion of poor households rent their dwelling units.

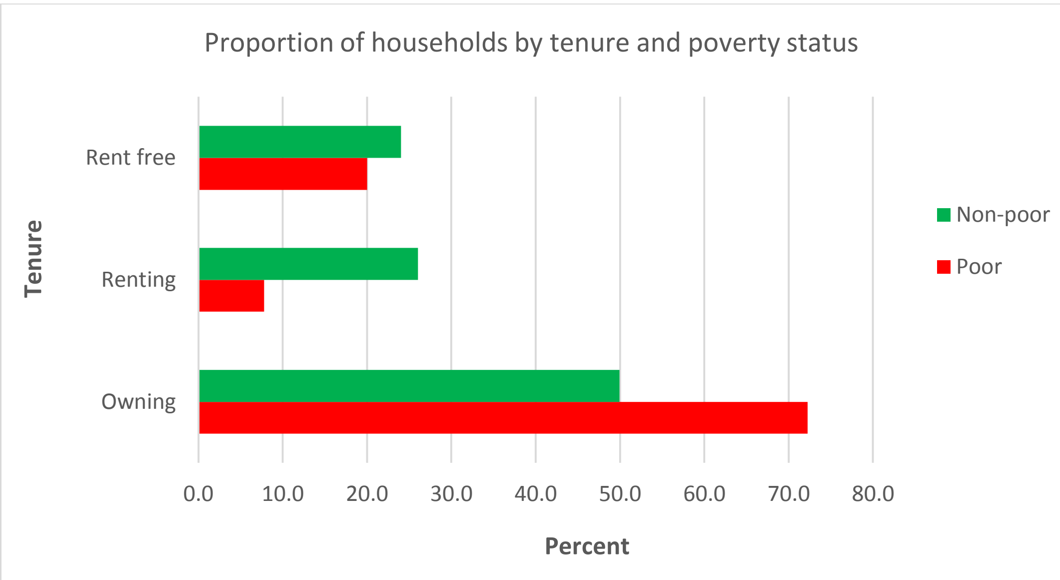

Figure 4.12: Proportion of households by tenure and poverty status

4.1.4 Household Assets Ownership and Poverty Status

The list of household assets common to both the GLSS6 and MICS4 data sets are used. These include land, refrigerator, television, laptop, desktop computer, motor cycle, bicycle, radio, CD player and washing machine. The individual assets, as variables in the data set, are grouped into whether or not households own any of them. These individual variables are further grouped into one variable known as "ownasset".

.

vi. Proportion of households by asset ownership and poverty status

In this section, the distribution of whether or not households own any asset over poverty status is shown in Table 4.1 and the distribution of ownership of individual asset over poverty status is displayed in Figure 4.13.

Table 4.1 indicates that high proportions of both the poor (84.8%) and non-poor households (92.8%) own assets. A very small proportion of the non-poor households (7.2%) do not any of the assets. Table 4.1, indicated that 90.9 percent of all households, both poor and non-poor, own at least one of the households the assets listed.

Figure 4.13 shows how many of the households own a particular asset and their poverty status. From the figure, the proportions of non-poor households who do not own any of these assets are far less than the poor households. Higher proportions of the poor households own bicycle, radio, land and motorcycle while higher proportions of the non-poor households own television, refrigerator and land. It can be observed that about the same proportion of poor and non-poor households own land.

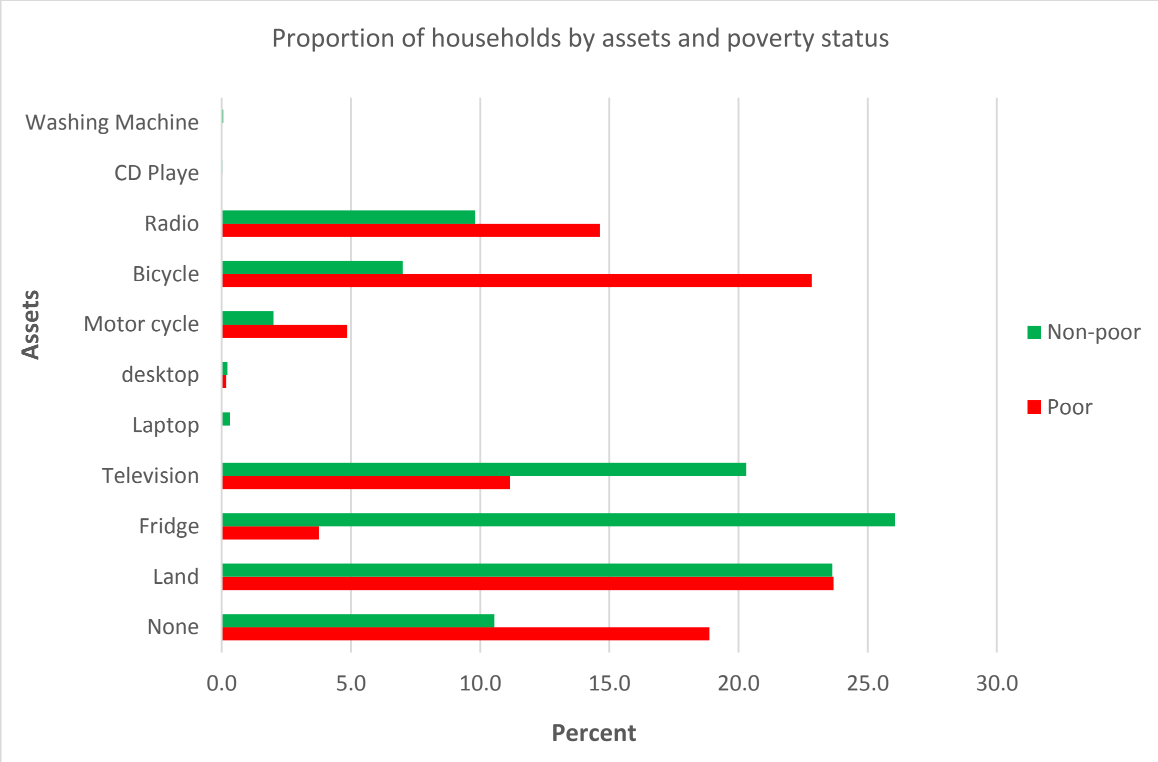

Figure 4.13: Proportion of households by particular asset owned and poverty status

4.2 REGRESSION MODEL ESTIMATION

The estimated regression model is of the form:

$$lny_h = \alpha + b_1x_{1h} + b_2x_{2h} + \cdots + b_kx_{kh} + \varepsilon_i.$$

Where h = 1, 2, ……, 16,693 and i = 1, 2, ……, 45.

$\boldsymbol{lny_h}$ = the natural logarithm of household expenditure per capita;

$\boldsymbol{\alpha}$ = the constant term;

$\boldsymbol{b_i}$ = estimated coefficient of household characteristic i;

$\boldsymbol{x_{ih}}$ = the independent regressors of household characteristic i of household h; and

$\boldsymbol{\mathcal{E}_i}$ = the error term of household characteristic i.

The specific estimate of each of the 45 parameters is indicated in Table 4.2.

The output of the stepwise regression technique, assuming a significance level (i.e. maximum allowable error) of 0.05 for entry of a variable into the model and 0.05001 for removal of a variable from the model, is presented in Table 4.2. Also, other corresponding statistics are presented below the table.

The number of observations reduced from 16,772 to 16,693 because cases with missing responses for certain variables were excluded from the analysis. The Root Mean Square Error (Root MSE = 0.51267) is the standard deviation of the regression and the closer to zero the better. The R-square shows the amount of variability in the response variable (Natural log of household expenditure) that is explained by the predictor variables (household characteristics). In this analysis, the R-square (53.8%) is almost the same as the Adjusted R-square (53.6%) since the number of predictor variables is relatively small (41) and the number of observations is very large (16,693). The value obtained for the adjusted R-square indicates a plausible association between household expenditure and the selected household characteristics in the model. Each t-value in column 4 of the table above tests the null hypothesis that the coefficient of a selected predictor is not different from 0. The two-tailed p-values in column 5 also test the same hypothesis that each coefficient is not different from 0. To reject this hypothesis, the p-value has to be less than significance level of 0.05.

From the table, all the p-values of the selected predictors are less than 0.05; therefore, the predictors have significant relationship with household expenditure. Also, the lack-of-fit (overall) test criteria provided (i.e. Prob. > F = 0.001 and F (40, 16652) = 266.83) indicate clearly that the model significantly fit the data. In other words, the estimated model performs very well with the sample (data).

Table 4.2: Estimated parameters of the stepwise regression model

b_i	lnrpcexp (dependent variable)	Coefficient	Standard Error	t	P > | t |	[95% Confidence	Interval]
	Location						
b_1	urban household	0.0362	0.0114	3.18	0.001	0.0139	0.0586
b_2	Central	-0.0404	0.0151	-2.67	0.008	-0.07	-0.0108
b_3	Volta	-0.1159	0.0162	-7.14	0.000	-0.1477	-0.0841
b_4	Eastern	-0.0696	0.0142	-4.89	0.000	-0.0975	-0.0417
b_5	Brong Ahafo	-0.0844	0.0152	-5.57	0.000	-0.1141	-0.0547
b_6	Northern	-0.3105	0.0183	-16.99	0.000	-0.3463	-0.2747
b_7	Upper East	-0.3283	0.0243	-13.53	0.000	-0.3759	-0.2808
b_8	Upper West	-0.6255	0.0271	-23.11	0.000	-0.6786	-0.5725
	Demographic						
b_9	Household Size	-0.0824	0.0017	-47.38	0.000	-0.0858	-0.079
b_{10}	Male head	-0.063	0.0103	-6.12	0.000	-0.0832	-0.0428
b_{11}	Dependency ratio	-0.1981	0.0188	-10.56	0.000	-0.2349	-0.1613
	Education						
b_{12}	Head still schooling	0.3569	0.039	9.15	0.000	0.2804	0.4333
b_{13}	Head schooled in the past	0.122	0.0187	6.54	0.000	0.0855	0.1586
b_{14}	Less than primary level	-0.0802	0.02	-4	0.000	-0.1194	-0.0409
b_{15}	Primary level	-0.0919	0.0196	-4.69	0.000	-0.1304	-0.0535
b_{16}	JSS/JHS level	-0.0717	0.0152	-4.71	0.000	-0.1015	-0.0419
b_{17}	Middle School level	-0.0766	0.0137	-5.59	0.000	-0.1035	-0.0497
b_{18}	Head can write	0.0412	0.0129	3.19	0.001	0.0158	0.0665
	Assets						
b_{19}	Household owns radio	0.026	0.0084	3.04	0.002	0.0091	0.0421
b_{20}	Household owns bicycle	0.096	0.0107	8.95	0.000	0.0746	0.1164
b_{21}	Household owns laptop	0.165	0.0188	8.79	0.000	0.1284	0.2022
b_{22}	Television	0.125	0.0116	10.8	0.000	0.1026	0.1481
b_{23}	Household own desktop	0.076	0.0171	4.44	0.000	0.0423	0.1091
b_{24}	Motor cycle	0.233	0.0141	16.56	0.000	0.2052	0.2603
b_{25}	Washing machine	0.168	0.0407	4.13	0.000	0.0884	0.2481
b_{26}	Household owns car	0.351	0.0182	19.36	0.000	0.3158	0.387
b_{27}	Household owns land	0.166	0.0096	17.3	0.000	0.1471	0.1847
b_{28}	Household owns fridge	0.074	0.0113	6.54	0.000	0.0518	0.0961

Table 4.2: Estimated parameters of the stepwise regression model (continued)

b_i	Lnrpcexp (dependent variable)	Coefficient	Standard Error	t	P > | t |	[95% Confidence Interval]	
	Dwelling						
b_{29}	Sleeping rooms	0.041	0.003	14.19	0.000	0.035	0.047
b_{30}	Renting dwelling	0.034	0.011	3.05	0.002	0.012	0.055
b_{31}	water_drinking2	-0.062	0.010	-6.24	0.000	-0.082	-0.043
b_{32}	water_drinking3	0.090	0.012	7.54	0.000	0.066	0.113
b_{33}	lighting3	-0.053	0.012	-4.49	0.000	-0.077	-0.030
b_{34}	fuel2	0.127	0.013	10.07	0.000	0.102	0.152
b_{35}	fuel3	0.279	0.017	16.12	0.000	0.245	0.313
b_{36}	fuel4	0.192	0.030	6.49	0.000	0.134	0.249
b_{37}	toilet2	0.133	0.020	6.79	0.000	0.095	0.172
b_{38}	toilet3	0.037	0.014	2.56	0.010	0.009	0.065
b_{39}	toilet4	0.089	0.017	5.21	0.000	0.055	0.122
b_{40}	toilet5	0.075	0.013	5.67	0.000	0.049	0.101
b_{41}	wall2	0.073	0.027	2.7	0.007	0.020	0.127
b_{42}	wall3	0.052	0.011	4.5	0.000	0.029	0.074
b_{43}	floor2	0.042	0.015	2.72	0.007	0.012	0.072
b_{44}	floor3	0.098	0.025	3.97	0.000	0.049	0.146
b_{45}	roof2	0.077	0.013	6.04	0.000	0.052	0.102
α	***Constant***	7.729	0.026	292.56	0.000	7.677	7.780

****Note: Variable names are described in Table 4.8 in appendix A.***

Number of observation	16693	R-squared	0.5376
F(44, 16648)	439.89	Adjusted R-squared	0.5364
Prob > F	0.0000	Root MSE	0.51267

4.3 TESTS OF MODEL ASSUMPTIONS

The tests of the model assumptions is performed on normality of the error terms, presence of heteroskedasticity, independence of the error term, linearity of the regression function and test for omitted variables bias. The test statistics for Lack-of-fit test are provided alongside the parameter estimates.

4.3.1 Test for Normality of Error Term

A histogram of the residuals is used to test for normality of the error term. Figure 4.14 shows that the histogram of the residuals is symmetrical and the mean is 0, therefore, the values of the error term are normally distributed.

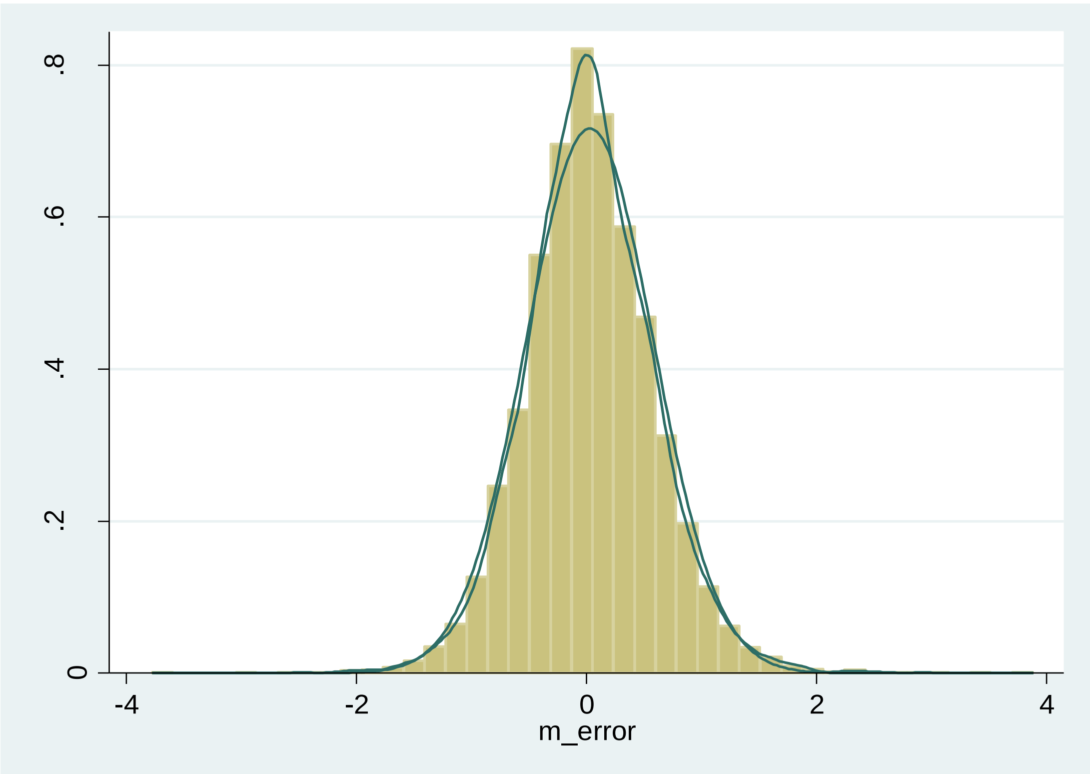

Figure 4.14: A histogram showing the distribution of the residuals.

4.3.2 Test for heteroskedasticity of error term

The outcome of test for normality of the error term above satisfies the condition to carry out the Breusch-Pagan / Cook-Weisberg test for heteroskedasticity. The results of the analysis provided the value of the test statistics and p-value ch2 (1) = χ^2_{bp} = 230.27 and Prob > chi2 = 0.001 respectively.

Since the p-value = 0.001 < 0.05 we fail to reject the null hypothesis and conclude that the variance of the error term is constant.

4.3.3 Test for independence of the error term

The requirements for Durbin-Watson test haven been satisfied by the preceding tests performed.

The results of the analysis provided Durbin-Watson d-statistic (45, 16693) = 1.967788

At 5% significance level, the lower bound (d_l) is 1.284 and the upper bound (d_u) is 1.567.

Comparing the d-statistic and the lower and upper bounds, Durbin-Watson d-statistic D = 1.968 > d_u = 1.567. Clearly, we do not reject the null hypothesis and conclude that there is no first order autocorrelation and that the predictor variables are strictly exogenous. This confirms that the error term of the estimated model is independent.

4.3.4 Test for linearity of the regression function

The Random-Versus-Fitted plot (rvfplot) is used to test for linearity of the regression function. Also a scatter plot of observe responses (lnrpcexp) and predicted responses (yhat) is used to confirm linearity of the regression function.

Figure 5.2 displays the random-versus-fitted plot while Figure 5.3 depicts the scatter plot of lnrpcexp and yhat. Figure 4.15 shows a linear pattern with most of the residual values concentrated between -2 and 2. Therefore there is a linear relationship between household expenditure and household characteristics. Furthermore, Figure 4.16 shows a linear pattern between the values of the observed and predicted responses, hence the regression function estimated is linear.

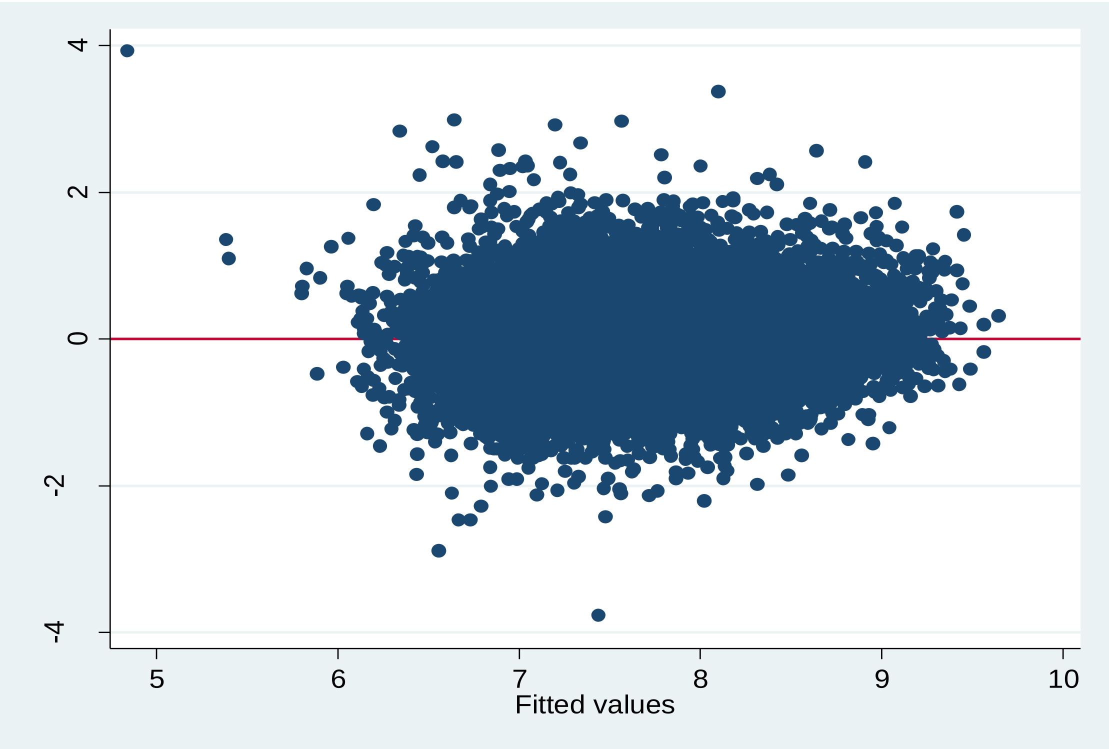

Figure 4.15: Random-Versus-Residual Plot showing linearity of the estimated model

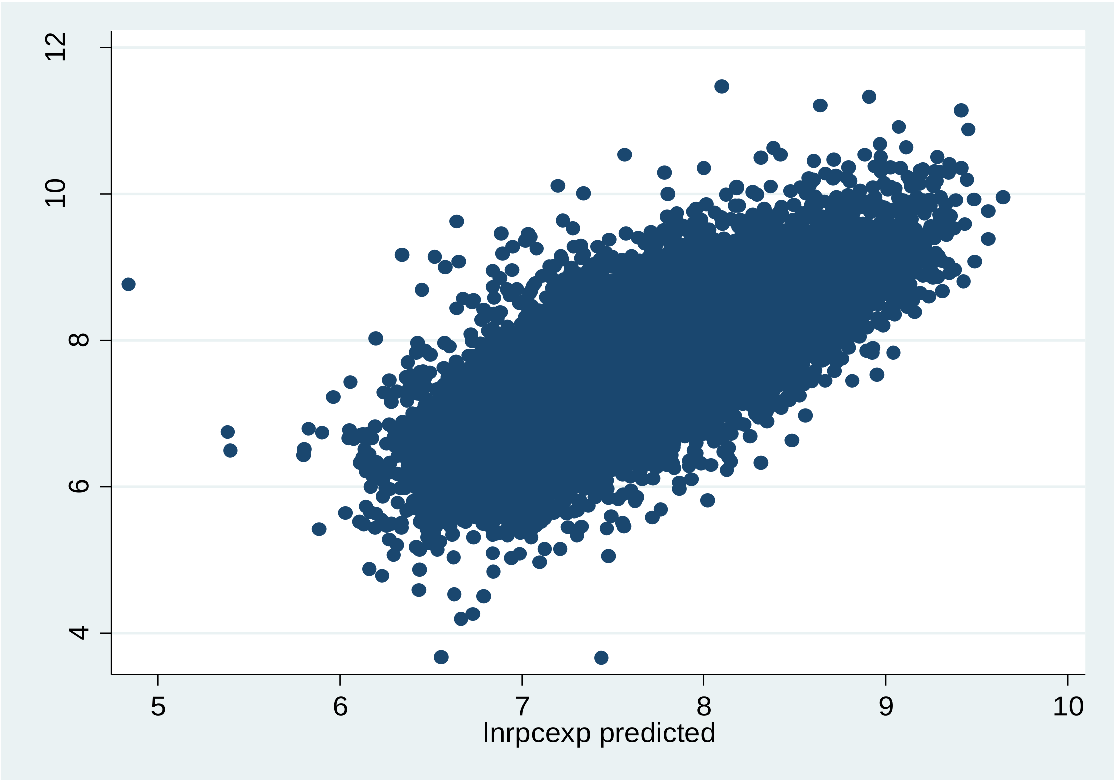

Figure 4.16: Scatter Plot of observed and predicted responses

4.3.5 Test for Omitted-Variable Bias

The purpose of this analysis is to determine whether there are any other key variables that could provide important additional descriptive and predictive power to the model. Model misspecification due to omission of important predictor variables tends to be serious, leading to biased estimates of the regression parameters and error variance. (Kutner et al, 2005).

The test statistic provided by running "ovtest" command in STATA version 12 is as follows. $F(3, 16645) = 69.47$ Prob. $> F = 0.0001$

The decision is that if p-value is greater than the significance level then do not reject the null hypothesis. From the statistics provided p-value = 0.001 < 0.05 therefore we reject the null hypothesis that the model has not omitted important variables and conclude that there important variables omitted from the estimated model above.

The outcome of this test does not imply that the model is inappropriate or misspecified, but it suggests absence of other predictor variables which significantly influence poverty (e.g. health, 'susu'- small scale savings and employment variables). However, this study is limited to variables of the same characteristics in two data sets, GLSS6 and MICS4. Thus, those variables absent cannot be included currently since they are not available.

Furthermore, according to Kutner et al, 2005, only a few of the factors operating on any response variable Y in real-world situations can be included explicitly in a regression model. The chief purpose of residual analysis in identifying other important predictor variables is therefore to test the adequacy of the model and see whether it could be improved materially by adding one or more predictor variables. Nevertheless, the Adjusted R-square (53.6%) indicates that the model estimated is adequate for further use.

4.4 MODEL VALIDATION

Validation is a useful and necessary part of the model-building process. It is essential to establish the stability and reasonableness of the regression model coefficients, the plausibility and usability of the regression function, and the ability to generalize inferences drawn from the regression analysis (Kutner et al, 2005). According to Kutner et al, 2005, the most preferred method to validate a regression model is through collection of new data set. However, this is neither practicable nor feasible in this study.

Therefore, this research uses a 10-fold cross-validation method to validate the model and obtain an optimum p-value for the final estimates, though some significant variables are not included in the model which is a limitation of this study. The results of the cross-validation are provided in Table 4.3.

Table 4.3 shows a predicted poverty headcount of 24.9 percent while the GLSS6 estimate is 24.2 percent. The model estimate also falls outside the confidence interval of the GLSS6 estimate. This can be attributed to the omitted variable bias resulting from the omission of employment and health variables as well as susu savings (small scale savings) that are known to have significant influence on household expenditure.

Table 4.3: Summary statistics from cross-validation of the model

Mean Estimates	Mean	Standard Error	[95% Confidence	Interval]
Predicted poor	0.249	0.0005	0.2488	0.2507
Poor (GLSS6 full data)	0.242	0.0009	0.2406	0.2443
Mean Absolute Difference	0.014	0.0007	0.0122	0.0148

4.5 ESTIMATING POVERTY INCIDENCE WITH THE MODEL USING MICS4 DATA SET

The regression model built using GLSS6 is applied to the MICS4 data set to impute annual expenditure values for the households in MICS4 using regression multiple imputation approach. This is made possible based on adequate predictor variables of similar characteristics identified in both surveys, though other variables present in GLSS6, which significantly influences poverty are not available in MICS4 and therefore not included in the model.

This section compares incidence of poverty estimated between GLSS6 and MICS4 data sets after application of the model to impute household expenditure into the MICS4 data set. Table 4.4 displays the comparison of estimates.

Table 4.4 shows that GLSS6 estimated a national poverty incidence to be 24.2 percent whereas the model predicted national poverty incidence to be 24.0 percent. Clearly, the estimate of poverty incidence predicted by the model is very close to the values obtained from GLSS6. Also, the poverty headcount estimate from the GLSS6 falls within the 95 percent confidence interval of the estimate from MICS4 data set which indicates that there is 95 percent probability that the model will predict poverty incidence accurately.

Table 4.4: Comparison of estimated poverty incidence from MICS4 with GLSS6

Mean Estimates	GLSS6 (2012/13)	MICS4 (2011)	Standard Error	[95% Confidence Interval]		Absolute Difference
Ghana	24.2	24.0	0.0060	0.2277	0.2512	0.20
Urban	10.6	8.4	0.0057	0.0726	0.0950	2.20
Rural	37.9	36.4	0.0097	0.3446	0.3827	1.50

This result indicates that incidence of poverty has worsened over the 2011 and 2012/2013 period.

4.6 SUMMARY OF RESULTS

The preliminary analysis reveals that higher proportions of households in the northern part of Ghana are poor as compared to those in the southern part. It is also observed there are more poor households in the rural than urban localities. There are more poor household heads in the working (or dependent) age group than those aged 65 year or more. The analysis indicated that households with small sizes are less likely to be poor whereas those with large sizes are more likely to become poor. It shows that household heads with higher level of education (secondary or higher) are less likely to be poor.

The analysis on dwelling facilities points out that higher proportions of non-poor households have access to safe drinking water. Also, higher proportions of poor households use less improved (i.e. wood, charcoal, was dust and animal waste) cooking fuel. Also, about 93.9 percent of the poor household do not use improved toilet facility or have no toilet facility. On main material for external wall, greater proportions of poor households' have nonconventional floor materials. Concerning main construction material of the floor, almost the same proportion of both the poor and non-poor households use cemented floor.

The analysis also shows that metal sheet is the main roofing material among the poor and non-poor. Higher proportions of both poor and non-poor households use metal sheet as the main material for roofing. It is observed that higher proportion of poor households is overcrowded, but they own their dwelling. Further investigations (Table 4.14 in Appendix A) revealed that most of these dwellings are huts and "single" rooms. The three northern regions of Ghana have been identified with high proportion of their population or households being poor. Most of these households, especially in the rural localities, live in huts or single rooms that are built by themselves. The analysis pointed out that about the same proportion (23.6%) of poor and non-poor households own land either for agriculture or commercial purpose.

All the assumptions underlying the model were satisfied and the household characteristics or variables in the model showed a significant relationship with the natural log of annual household expenditure. The overall fit was satisfied and the model was proven to be valid. The model estimated poverty incidence value (24.0%) that was very close to the value provided by the GLSS6 (24.2%). It was observed that the model has 95 percent probability of estimating poverty incidence accurately at the national level. At the urban and rural level the estimates were also close with absolute percentage difference of 2.2 and 1.5 respectively.

CHAPTER FIVE

DISCUSSION, CONCLUSION AND RECOMMENDATION

5.0 SUMMARY

The main purpose of this thesis is to develop a regression model from a household expenditure survey (Ghana Living Standards Survey) and use that model to estimate expenditure values for households in another data set which has no expenditure variable. This is researched in order to estimate incidence of poverty and other expenditure-based poverty indicators using other non-expenditure household surveys that are conducted in-between the Ghana Living Standards Surveys of which the data sets becomes available in seven years interval.

The thesis identified and obtained two data sets; the current household expenditure survey (GLSS6) and another (non-expenditure) household survey (MICS4) which were not too different in terms of time and other household characteristics. Common variables existing between the two data sets were identified. The GLSS6 data set was then then used to develop a stepwise regression model which selected the predictor variables (household characteristics) among the common variables that significantly relate to the household expenditure (response variable). A cross-validation method was employed to validate the performance of the model.

Applying the model to the MICS4 data set based on the selected common predictor variables and assuming that the household expenditure was completely missing at random in the MICS4 data set, multiple imputation techniques was used to impute the expenditure values for the households drawing on the observed values of the predictor variables and the prior distribution of the expenditure variable in the model.

The national poverty incidence estimated using this technique (24.0%) came close to that established by the sixth round of the Ghana Living Standards Survey (24.2%) with 0.2 percentage difference. The estimates for urban and rural localities were also not too different as shown in the analysis in chapter four.

5.1 DISCUSSION

The series of Ghana Living Standards Survey (GLSS), which are household income and expenditure surveys are the main source of data for estimating incidence of poverty and other poverty indicators. The indicators estimated are used to monitor effects of government programmes and projects towards alleviating poverty and improving the living conditions of the population. It is also used to track progress towards achieving the Millennium Development Goals (MDGs). However, the next GLSS data set becomes available after seven years since its inception. The poverty indicator estimated from a particular GLSS survey is used for the years in-between the surveys making the indicators constant over these years.

This thesis sets out to examine the application of stepwise regression model and multiple imputation technique to estimate poverty indicators from other non-expenditure household surveys that are usually conducted in-between the living standard surveys. This is purported to provide proxy measures of poverty incidence based on imputed household expenditure, and also, to help monitor and make short term adjustments to programmes and policies towards the poor.

Rubin (1976) developed this inferential methodology for missing data. The author proposed imputing missing values multiple times to have multiple data sets, and perform analysis of each data set to have multiple estimates. He further suggested combining the parameters for all analysis to have a single point estimate. The estimates are then combined to reflect within imputation and between imputation variability (Rubin, 1976; Marwala, 2009). The Expectation Maximization (EM) algorithm was also developed by Dempster, Laird, and Rubin (1977) for missing data estimation. Since then, researchers have been applying different methods to analyse missing data in different circumstances such as case deletion, pairwise deletion, simple-rule prediction, mean substitution, hot-deck imputation, cold-deck imputation, imputation using regression, regression-based nearest neighbor hot-decking, tree-based imputation, and stochastic imputation (Marwala, 2009). Little and Rubin (1987) identified some issues regarding case deletion and the single imputation methods. They identified that case deletion methods may reduce statistical power, and single imputation (mean imputation) may result in underestimating the variance of estimates. As a result, Rubin (1987) developed the multiple imputation method to deal with missing data analysis without these issues (Garg, 2013).

Graham (2009) established three criteria for a good missing data method. First, the method should yield unbiased estimates for a variety of parameters. Second, the method should include a way to assess uncertainty due to the missing observations. Third, the method should have good statistical power. Multiple imputation fulfills these criteria, especially when the missing observations are Missing Completely at Random (MCAR).

The application of multiple imputations in this thesis is quite different from very many of its use. Most researchers use multiple imputations to make incomplete data set complete for 'proper' analysis by estimating the missing values for one or more variables. Others also compare different methods of handling missing values and identify multiple imputations to be the superior. This thesis rather applies multiple imputations to estimate values of a particular variable for each observation based on the assumption that all the values are missing for that variable.

A study conducted by Garg (2013) with simulation based on a monotonic missing data pattern for sample sizes of 100, 500, 1000 and 5000 and missing data ranging from 10% to 25% with increments of 5% indicated that multiple imputation works well when compared to other single imputation methods such as mean substitution and single regression imputation. The study again showed that multiple imputations performed well against the available data analysis. Furthermore, Priyanka Garg recommended that multiple imputations is an effective technique to deal with missing data in public health research and clinical trials because it provides valid inferences.

It is worth considering the limitation of this study. Some of the household variables in the GLSS6 data set that are empirically proven to largely explain variability in household expenditure were not identified in the MICS4 data set. For example, employment and 'susu' (i.e. small savings mechanism) could not be found in the MICS4 data set. Also, according to Schafer and Graham, (2002) ignoring the distinct characteristics of the households in the two data sets, when constructing the imputation model, fails to preserve the relationships and characteristics of the data set which is the main goal of the imputations. Further, MICS4 was the nearest data set to the GLSS6 which have one and half year difference in the time the data was collected.

The study continued with the available common predictor variables since they explained about 54 percent of the total variability in household expenditure. The overall model was significant in explaining household expenditure. With the multiple imputation regression technique, the estimated incidence of poverty using the MICS4 data set came close to that of the already established poverty incidence from the GLSS6 for national, and urban and rural levels.

5.2 CONCLUSIONS

The methodology as well as the model is developed, and now, it is possible to estimate, using a household expenditure survey as a baseline, poverty indicators from a non-expenditure household survey to increase the frequency and availability of these indicators for policy adjustments and redirection. From the analysis poverty in Ghana still remains a rural phenomenon. There is higher proportion of rural households being poor than that of the urban. In terms of age, higher proportion of household heads in the dependent age group (15-64 years) is poor. The analysis also indicated that households with small sizes are less likely to be poor whereas those with large sizes are more likely to be poor. In addition, household heads with higher level of education are less likely to be poor and the opposite is shown to be true.

The analysis on dwelling characteristics pointed out that higher proportion of the non-poor households live in conventional dwellings with quality materials for the wall, floor and roofing; have access to many rooms for sleeping; have access to improved source of drinking water and toilet facilities; have improved source of lighting and fuel for cooking; and ownership of household assets including land for commercial or agriculture purpose. The reverse was the situation for the poor households.

With the exception of the omitted variable bias which showed that some significant predictor variables are not included in the regression model, all the other assumptions underlying the model as well as multiple imputations were satisfied. The omitted variables could not be included since they were not common to both data sets. The methodology (model) was able to estimate the national poverty incidence to be 24.0 percent from the MICS4 data set which was very close to the 24.2 percent established by the GLSS6. The estimation was done at a 95 percent probability. Certainly, the values are not expected to be that same since there is an interval of about two years between the two data sets.

5.3 RECOMMENDATIONS

The model developed aimed at estimating poverty from a non-expenditure household survey with Living Standards Survey as the baseline. This idea was conceived due to the paucity of household expenditure survey data sets to provide expenditure based poverty indicators frequently for programmes and project monitoring and adjustments. The methodology has by this thesis shown to be effective in estimating poverty incidence accurately at 95 percent probability. This study therefore recommends the following.

i. This approach or methodology should be considered, improved if possible, and utilized to increase the frequency of poverty indicators.

ii. Non-expenditure household surveys with relatively simple questionnaire, unlike that of the GLSS, should be conducted biennially to cover all household characteristics, including those omitted by this model, identified to significantly explain variability in household expenditure and apply this method to estimate poverty indicators.

iii. Though national poverty incidence has fallen consistently, poverty incidence in the rural localities of Ghana is still high. This study recommends two-prong approach to poverty reduction programmes and projects: rural poverty reduction programme and urban poverty reduction programme with significant difference between them. This will help speed up poverty reduction in the rural communities and eventually further reduce the national poverty incidence.

iv. Education provision should be standardized between the private and public schools as well as between rural and urban areas. Knowledge and skills acquisition and application should be encouraged among the poor, especially those in the rural localities.

v. Knowledge and skills provision should not be limited only to formal education. The functional literacy programme should be redesigned to include skills training and entrepreneurship for the rural poor.

REFERENCES

Alkire, S., (2007). "*The missing dimensions of poverty data: Introduction to the special issue*." Oxford Development Studies 35(4): 347 - 359.

Alkire, S. & Foster, J., E., (2007). *Counting and multidimensional poverty measures*: Working Paper 7, Oxford Poverty and Human Development Initiative, University of Oxford.

Alkire, S. (2014). *Towards Frequent and Accurate Poverty Data*: Oxford Poverty and Human Development Initiative (OPHI), Briefing Paper.

Barnard, J. & Rubin, D.B. (1999). *Small-Sample Degrees of Freedom with Multiple Imputation*, *Biometrika*, 86, 948–955

Berthoud, R., & Bryan, M., (2011). Income, Deprivation and Poverty: A Longitudinal Analysis. *Journal of Social Policy*, 40, pp 135-156. DOI: .1017/S0047279410000504.

Chaudhuri, S., Jalan, J., & Suryahadi, A. (2002). *Assessing household vulnerability to poverty from cross-sectional data: A methodology and estimates from Indonesia:* Columbia University, Department of Economics; Discussion Paper #:0102-52.

Dempster, A. P., Laird, N., M., & Rubin, D. B., (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of Royal Statistical Society*, 39, 1-38.

Elbers, C., Lanjouw, J., O., & Lanjouw, P., (2002). *Micro-Level Estimation of Welfare*: World Bank Policy Research Working Paper No. WPS 2911

Elbers, C., Lanjouw, J., O., & Lanjouw, P., (2003). *Micro-level Estimation of Poverty and Inequality*: *Econometrica*, 71(1):355–364.

Filmer, D. & Pritchett, L. H. (2001). *Estimating wealth effects without expenditure data or tears: An application to educational enrollments in states of India.* Demography, 38(1), 115-132.

Foster, J. E., Greer J. & Thorbecke E. (1984). A class of decomposable poverty measures: *Econometrica*, 56(1), 173-177.

Garg, P., (2013). *Robustness of multiple imputation under missing at random (MAR) mechanism:* A simulation study.

Ghana Statistical Service (1999). *Poverty Trends in the 1990's,* Ghana Statistical Service Report.

Ghana Statistical Service (2000). *Ghana Living Standards Surveys:* Report on the fourth round (GLSS4).

Ghana Statistical Service (2003). *Ghana Demographic and Health Survey Report.*

Ghana Statistical Service (2007). *Patterns and Trends of Poverty in Ghana Report.*

Ghana Statistical Service (2007). *Ghana Living Standards Surveys:* Report on the fifth round (GLSS5).

Ghana Statistical Service (2013). *Non-Monetary Poverty in Ghana*: Population and Housing Census (2010 PHC).

Ghana Statistical Service, (2013). *National Analytical Report (2010):* Population and Housing Census (2010 PHC).

Ghana Statistical Service, (2014). '*Poverty Profile in Ghana (2005-2013):* Ghana Living Standards Survey round six' (GLSS6)

Glewwe, P. (1991). *Investigating the determinants of household welfare in Cote d'ivoire, living standards measurement study: Working Paper No. 71*. Washington DC: The World Bank.

Glewwe, P. (2001). Trade liberalization and poverty: the evidence so far. *Journal of Economic Literature*, 42(1), 72-115.

Goldstein, R. (1991). srd5: Ramsey test for heteroscedasticity and omitted variables. *Stata Technical Bulletin 2: 27. Reprinted in Stata Technical Bulletin Reprints, vol. 1, p. 177*. College Station, TX: Stata Press.

Gönner, C., Haug, M., Cahyat, A., Wollenberg, E., De Jong, W., Limberg, G., Cronkleton, P., Moeliono, M., & Becker, M., (2010). *Capturing Nested Spheres of Poverty*: A model for multidimensional poverty analysis and monitoring; CIFOR Occasional Paper No. 46

Graham, J., W., (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549-576.

Heitjan D., F., & Little R., J., A., (1991). Multiple imputation for the fatal accident reporting system. *Journal of the Royal Statistical Society* C, 40(1), 13–29.

Hornblad, J., (2013). Missing data in the Swedish national patients register: *Multiple imputation by fully conditional specification*. Department of Statistics, Stockholm University.

Horton, N. J., & Lipsitz, S. R. (2001). *Multiple imputation in practice: Comparison of software packages for regression models with missing variables.* The American statistician, 55, 244–254.

Kutner, M., H., Nachtsheim, C., J., Neter, J., & Li, W., (2005). *Applied linear statistical models;* 5th ed. p. cm. - (McGraw-Hill/Irwin series Operations and decision sciences)

Lanjouw, J., & Lanjouw, P., (2001). How to compare apples and oranges: Poverty measurement based on different definitions of consumption. *Review of income and wealth*, 47(1), 25-42.

Lavori, P., W., Dawson, R., & Shera, D., (1995). *A multiple imputation strategy for clinical trials with truncation of patient data*. Statistics in medicine, 14, 1913-1925. http://dx.doi.org/10.1002/sim.4780141707

Little, R.J.A. & Rubin, D.B. (1987), *Statistical Analysis with Missing Data*, New York: John Wiley & Sons, Inc.

Mack, J., & S., Lansley (1985). '*Poor Britain'*. London, George Allen & Unwin.

Maliki, S. (2011). Monetary and non-monetary deprivation in Algeria: *A determinant method by the ROC curve. A paper presented at the Special IARIW-SSA Conference on Measuring National Income, Wealth, Poverty and Inequality in African Countries*; Cape Town, September 28-October 1, 2011.

Marwala, T., (2009). Introduction to Missing Data. In Marwala, T. (Ed.), *Computational Intelligence for Missing Data Imputation, Estimation, and Management: Knowledge Optimization Techniques.* (pp. 1-18). doi:10.4018/978-1-60566-336-4.ch001.

Maxwell, S., (1999). *The Meaning and Measurement of Poverty*: ODI Poverty Briefing, 3rd Edition

Meng, X., & Gregory, R., (2007). Urban Poverty in China and Its Contributing Factors, 1986-2000. *Review of Income and Wealth, 53(1),* 167-189.

Nolan, B., & Whelan, C., T., (2010). Using non-monetary deprivation indicators to analyze poverty and social exclusion: Lessons from Europe: *Journal of Policy Analysis and Management* 29(2), pp. 305-325.

Osberg, L. & Sharpe, A., (2005). How should we measure the "economic" aspects of well-being: *Review of Income and Wealth* 51(2), pp. 311-336.

Ravallion, M., (1992). Poverty Comparisons: *A guide to concepts and methods.* Living standards measurement study: Working Paper No. 88, World Bank, Washington, D. C

Ravallion, Martin (2002). *How Not to Count the Poor*: A reply to Reddy and Pogge. World Bank working paper.

Rencher, C., Alvin (2002). *Methods_of_multivariate_analysis*; 2nd. ed. p. cm. (Wiley series in probability and mathematical statistics) "A Wiley-Interscience publication

Rosenbaum, P., R. & Rubin, D., B., (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, Vol. 70, No. 1. (Apr., 1983), pp. 41-55

Rowntree, B., S., & Lavers, G., R., (1951). *Poverty and the Welfare State* (London: Longmans).

Rubin, D., B. (1976). Inference and Missing Data. *Biometrika, 63, 581-592*.

Rubin, D., B. (1987). *Multiple Imputation for Nonresonse in Surveys.* NewYork. Wiley

Rubin, D., B. (1996). Multiple Imputation after 18+ years. *Journal of American Statistical Society, 91, 473-489.*

Sahn, D., & Stifel, D., (2000). Poverty comparisons over time and across countries in Africa, *World Development*, 28-12: 2123-55.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data. London:* Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing Data: Our View of the State of Art. *Psychological Methods*, 7, 147-177.

Schafer, J. L. (2003). Multiple imputation in multivariate problems: when the imputation and analysis models Di_er. *Statistica Neerlandica*, 57, 19-35.

Schafer, J.L. (1999), NORM*: Multiple Imputation of Incomplete Multivariate Data under a normal model, version 2*.

Schenker, N., & Taylor J., M., G., (1996). *Partially Parametric Techniques for Multiple Imputation*. Computational Statistics & Data Analysis, 22(4), 425–446.

Sen, A. (1985). Well-being, agency and freedom: the Dewey lectures 1984: *The Journal of Philosophy*, 82(4), pp. 169-221.

Sen, A. (1993). Capability and well-being, M. Nussbaum & A. Sen (Eds.): *The Quality of Life*. Oxford: Clarendon Press, pp. 30-53.

Sen, A. (2000). A decade of human development: *Journal of human development* 1(1), pp. 17-23.

Stein R. (1988). Direct and Indirect Measures of Poverty. *Journal of social policy*, 17, pp 351-365. DOI: 10.1017/S0047279400016858.

Stock, J., H. & Watson M., W., (2003a): *Introduction to Econometrics*. Boston: Addison Wesley Longman. *Fed. Reserve Bank Richmond Econ. Quart.* 89:3.

Todaro, M., P., & Smith S., C., *Economic Development*, 2012– 11th edition

Torres-Reyna, O., (2007). *Linear regression using stata (v.6.3);* Princeton University (available at: http://www.princeton.edu/~otorres/Regression101.pdf), 5th June, 2015.

Townsend, P. (2002). Country: Peasant household behavior with missing markets: some paradoxes explained. *The Economic Journal*, *101*(409): 1400-1417.

Townsend, P. (2010). The meaning of poverty: *British Journal of Sociology* DOI: 10.1111/j.1468-4446.2009.01241.x, pp. 85-102.

UNDP 2010: *Human Development Report 2010*, New York: UNDP

UNSD (2005). *Special Project on Poverty Statistics:* Handbook on Poverty Statistics: Concepts, Methods and Policy Use (*available at: https://www.google.com.gh/?gws_rd=cr,ssl&ei=0MWoVef0NITD7gaRg67oCg#q=Handbook+on+Poverty+Statistics:+Concepts%2C+Methods+and+Policy+Use), 9th November 2014.*

Van, Buuren, S. (2012*). Flexible Imputation of Missing Data*. Chapman & Hall/CRC

World Bank (1995). Ghana: poverty, past, present and future, report 14504 GH

World Bank (2000). *World development report 2000/2001:* Attacking poverty

World Bank (2005). *Introduction to poverty analysis:* World Bank institute; August, 2005

World Bank (2007). *World Development Indicators.*

World Health Organization (1948). Constitution of the World Health Organization. *American journal of public health and the nations' health*: November 1946, Vol. 36, No. 11, pp. 1315-1323.DOI:10.2105/AJPH.36.11.1315, (available at: http://www.who.int/ governance/eb/who_constitution_en.pdf)

Yang C., Y., (2011). Multiple imputation using SAS software: *Journal of statistical software;* December 2011, Volume 45, Issue 6 (available at: http://www.jstatsoft.org).

APPENDIX A

Table 4.2a: Distribution of poverty incidence by region, 1991/92 – 2012/13

Region	1991-92	1998-99	2005-06	2012-13
Ghana	51.7	39.5	28.5	24.2
Western	59.6	27.3	18.4	20.9
Central	44.3	48.4	19.9	18.8
Greater Accra	25.8	5.2	11.8	5.6
Volta	57.0	37.7	31.4	33.8
Eastern	48.0	43.7	15.1	21.7
Ashanti	41.2	27.7	20.3	14.8
Brong Ahafo	65.0	35.8	29.5	27.9
Northern	63.4	69.2	52.3	50.4
Upper East	66.9	88.2	70.4	44.4
Upper West	88.4	83.9	87.9	70.7

Source: Ghana Statistical Service, Ghana Living Standards (1991/92, 1998/99, 2005/06, 2012/13)

Table 4.3a: Distribution of Poverty Gap Ratio and Poverty Incidence by year

Years	Poverty Gap Ratio	Poverty Incidence
1991-92	18.5	51.7
1998-99	13.9	39.5
2005-06	9.6	28.5
2012-13	7.8	24.2

Source: Ghana Statistical Service, Ghana Living Standards (1991/92, 1998/99, 2005/06, 2012/13)

Table 4.4a: Distribution of poverty gap ratio by region

Region	1991-92	1998-99	2005-06	2012-13
Ghana	18.5	13.9	9.6	7.8
Western	20.5	7.0	4.2	5.7
Central	12.9	14.8	4.3	5.6
Greater Accra	6.3	1.1	3.1	1.6
Volta	20.1	9.9	7.3	9.8
Eastern	15.9	15.6	3.3	5.8
Ashanti	12.9	8.5	5.2	3.5
Brong Ahafo	22.8	9.8	7.8	7.4
Northern	29.9	29.9	20.7	19.3
Upper East	28.7	44.0	32.7	17.2
Upper West	41.3	38.8	48.0	33.2

Source: Ghana Statistical Service, Ghana Living Standards (1991/92, 1998/99, 2005/06, 2012/13)

Table 4.5a: Distribution of poverty incidence by locality of residence and year

Year	Urban	Rural
1991/92	27.7	63.6
1998/99	19.4	49.5
2005/06	10.8	39.2
2012/13	10.6	37.9

Source: Ghana Statistical Service, Ghana Living Standards (1991/92, 1998/99, 2005/06, 2012/13)

Table 4.6a: Proportion of heads by highest level of education and poverty status

Highest Education Level	Poor	Non-poor	Total
None	51.2	20.3	27.8
Less than primary	11.2	8.8	9.4
Primary	8.5	7.6	7.8
JSS/JHS	10.6	16.5	15.1
Middle school	13.8	23.8	21.4
SSS/SHS or higher	4.6	23.1	18.6
Total	100.0	100.0	100.0

Source: Ghana Statistical Service, Ghana Living Standards, 2012/13

Table 4.7a: Proportion of households by source of main drinking water and poverty status of head

Main drinking water source	Poor	Non-poor	Total
Pipe inside or outside dwelling, public tap	18.6	31.3	28.2
Bore-hole/pump/tube well, protected well	52.5	28.4	34.3
Bottled or satchet water	3.4	29.1	22.8
Rain water, tanker, unprotected well or spring etc.	25.5	11.2	14.7
Total	100.0	100.0	100.0

Source: Ghana Statistical Service, Ghana Living Standards, 2012/13

Table 4.8a: Proportion of main cooking fuel of households by poverty status of head

Main source of cooking fuel	Poor	Non-poor	Total
Wood	83.9	39.2	50.1
Charcoal	13.7	34.7	29.6
Gas	1.1	23.7	18.2
Others	1.3	2.3	2.1
Total	100.0	100.0	100.0

Source: Ghana Statistical Service, Ghana Living Standards, 2012/13

Table 4.9a: Proportion of main external wall material of by poverty status of head

Main external wall material	Poor	Non-poor	Total
Mud bricks/earth, landcrete	66.5	26.7	36.4
Wood, metal sheet/slate/asbestos, bamboo	1.7	2.7	2.5
Stone, Burnt bricks, Cement blocks/concrete	31.8	70.6	61.2
Total	100.0	100.0	100.0

Source: Ghana Statistical Service, Ghana Living Standards, 2012/13

Table 4.10a: Distribution of main floor material by poverty status of head

Main floor material	Poor	Non-poor	Total
Earth/mud	16.8	5.9	8.5
Cement/concrete	81.0	88.2	86.5
Wood/vinyl/ceramic tiles	2.2	5.9	5.0
Total	100.0	100.0	100.0

Source: Ghana Statistical Service, Ghana Living Standards, 2012/13

Table 4.11a: Distribution of main roof material of household by poverty status of head

Main roof material	Poor	Non-poor	Total
Mud/mud bricks/earth, bamboo, palm leaves	21.2	6.7	10.3
Wood, slate/asbestos, roofing tile	7.3	13.0	11.7
Metal sheet	70.5	78.2	76.3
Concrete/Other	0.9	2.1	1.8
Total	100.0	100.0	100.0

Source: Ghana Statistical Service, Ghana Living Standards, 2012/13

Table 4.12a: Distribution of proportion by tenure and poverty status of head

Tenure	Poor	Non-poor	Total
Owning	72.2	49.9	55.4
Renting	7.8	26.1	21.6
Rent free	20.0	24.0	23.0
Total	100.0	100.0	100.0

Source: Ghana Statistical Service, Ghana Living Standards, 2012/13

Table 4.13a: Distribution of household assets owned by poverty status of head

Household assets owned	Poor	Non-poor	Total
None	18.9	10.6	12.5
Land	23.7	23.6	23.6
Fridge	3.8	26.1	20.7
Television	11.2	20.3	18.1
Laptop	0.0	0.3	0.3
desktop	0.2	0.2	0.2
Motor cycle	4.9	2.0	2.7
Bicycle	22.8	7.0	10.8
Radio	14.6	9.8	11.0
CD Player	0.0	0.0	0.0
Washing Machine	0.0	0.1	0.0
Total	100.0	100.0	100.0

Source: Ghana Statistical Service, Ghana Living Standards, 2012/13

Table 4.14: Proportion of dwelling type owned by poverty status of head of household

Type of Dwelling	Poor	Non-poor	Total
Separate house	11.9	17.9	16.5
Semi-detached house	8.2	7.0	7.3
Flat/Apartment	0.9	4.0	3.3
Compound house (rooms)	49.5	59.8	57.3
Huts/Buildings (same compound)	25.0	8.4	12.4
Huts/Buildings (different compounds)	2.1	0.9	1.2
Tent	0.1	0.0	0.0
Improvised home (kios	0.3	0.8	0.7
Living quarters attached	0.0	0.4	0.3
Uncompleted building	0.2	0.8	0.6
Other	1.8	0.1	0.5
Total	100.0	100.0	100.0

Source: Ghana Statistical Service, Ghana Living Standards, 2012/13

Table 4.7: Descriptive statistics on the selected variable

Variable	Observation	Mean	Std. Dev.	Minimum	Maximum
lnrpcexp	16772	7.75	0.829	3.664	11.48
urban	16772	0.44	0.497	0	1
region	16772	5.36	2.794	1	10
hhsize	16772	4.26	2.784	1	29
head_male	16772	0.72	0.450	0	1
head_age	16772	45.85	15.915	15	99
head_schooling	16772	2.35	0.925	1	3
schlvl	16766	3.32	1.945	1	6
read	16772	0.53	0.499	0	1
write	16772	0.51	0.500	0	1
rooms	16772	2.21	1.671	1	18
bedrooms	16772	1.90	1.366	1	32
tenure	16772	1.73	0.842	1	3
water_drinking	16772	2.27	1.015	1	4
fuel	16737	1.75	0.897	1	4
water_general	16772	1.86	0.749	1	3
wall	16772	2.11	0.981	1	3
toilet	16750	3.02	1.629	1	5
floor	16772	1.93	0.371	1	3
roof	16772	2.67	0.698	1	4
fridge	16766	0.28	0.448	0	1
radio	16766	0.54	0.498	0	1
desktop	16766	0.05	0.223	0	1
laptop	16766	0.06	0.235	0	1
tv	16766	0.50	0.500	0	1
bicycle	16766	0.30	0.460	0	1
mcycle	16766	0.11	0.310	0	1
land	16766	0.24	0.425	0	1
car	16766	0.05	0.209	0	1
washm	16766	0.01	0.087	0	1
cdplayer	16766	0.05	0.225	0	1

Table 4.8: Description of household variables common to both Surveys

Variable Name	Type	Display Format	Variable Label
rpcexp	float	%9.0g	Household expenditure
urban	float	%9.0g	Locality of residence: Urban = 1, Rural = 0
region	byte	%13.0g	Region
hhsize	byte	%8.0g	Household size
head_male	float	%9.0g	Household head is male = 1, 0 otherwise
head_age	float	%9.0g	Age of the household head
head_schooling	float	%27.0g	School attendance of the household head
schlvl	float	%26.0g	Highest level of schooling completed by the household head
read	float	%9.0g	Household head can read a simple sentence in any language
write	float	%9.0g	Household head can write a simple sentence in any language
rooms	float	%9.0g	Number of rooms
bedrooms	float	%9.0g	Number of bedrooms
tenure	byte	%30.0g	Tenancy arrangement
water_drinking	byte	%94.0g	Main source of drinking water
fuel	byte	%67.0g	Main source of cooking fuel
water_general	byte	%94.0g	Main source of water for general use
wall	byte	%75.0g	Main construction material of outer wall
toilet	byte	%32.0g	Type of toilet
floor	byte	%86.0g	Main construction material of floor
roof	byte	%63.0g	Main construction material of roof
fridge	float	%9.0g	Any household member or head owns a refrigerator
radio	float	%9.0g	Any household member or head owns a radio
desktop	float	%9.0g	Any household member or head owns a desktop computer
laptop	float	%9.0g	Any household member or head owns a laptop computer
tv	float	%9.0g	Any household member or head owns a television
bicycle	float	%9.0g	Any household member or head owns a bicycle
mcycle	float	%9.0g	Any household member or head owns a motor cycle
land	float	%9.0g	Any household member or head owns a land
car	float	%9.0g	Any household member or head owns a car
washm	float	%9.0g	Any household member or head owns a washing machine
cdplayer	float	%9.0g	Any household member or head owns a CD/DVD player

Table 5.1: Recoded Categorical Variables into Binary Variables

Location Variables		Education and Literacy Variables	
urban =	1, urban 0, Otherwise	Head_male =	1, head is a male 0, Otherwise
region1 =	1, Western 0, Otherwise	head_schooling1 =	1, Head never attended 0, Otherwise
region2 =	1, Central 0, Otherwise	head_schooling2 =	1, Head still attending school 0, Otherwise
region3 =	1, Greater Accra 0, Otherwise	head_schooling3 =	1, Head attended school in the past 0, Otherwise
region4 =	1, Volta 0, Otherwise	head_schlvl1 =	1, None 0, Otherwise
region5 =	1, Eastern 0, Otherwise	head_schlvl2 =	1, Less than primary 0, Otherwise
region6 =	1, Ashanti 0, Otherwise	head_schlvl3 =	1, Primary 0, Otherwise
region7 =	1, Brong Ahafo 0, Otherwise	head_schlvl4 =	1, JHS/JSS 0, Otherwise
region8 =	1, Northern 0, Otherwise	head_schlvl5 =	1, Middle School 0, Otherwise
region9 =	1, Upper East 0, Otherwise	head_schlvl6 =	1, SHS/SSS or higher 0, Otherwise
region10 =	1, Upper West 0, Otherwise	read =	1, Head can read 0, Otherwise
		write =	1, Head can write 0, Otherwise

Table 5.1: Recoded Categorical Variables into Binary Variables (continued)

Dwelling Characteristics	
tenure1 =	1, Owning 0, Otherwise
tenure2 =	1, Renting 0, Otherwise
tenure3 =	1, Rent free, perching etc. 0, Otherwise
water_drinking1 =	1, Pipe inside or outside dwelling, public tap 0, Otherwise
water_drinking2 =	1, Bore-hole/pump/tube well, protected well, protected spring 0, Otherwise
water_drinking3 =	1, Bottled or satchet water 0, Otherwise
water_drinking4 =	1, Rain water, tanker, unprotected well or spring etc. 0, Otherwise
water_general1 =	1, Pipe inside or outside dwelling, public tap 0, Otherwise
water_general2 =	1, Bore-hole/pump/tube well, protected well, protected spring, satchet water 0, Otherwise
water_general3 =	1, Rain water, tanker, unprotected well or spring, river etc. 0, Otherwise
fuel1 =	1, Wood 0, Otherwise
fuel2 =	1, Charcoal 0, Otherwise
fuel3 =	1, Gas 0, Otherwise
fuel4 =	1, Electricity, kerosene , crop residue etc. 0, Otherwise

Table 5.1: Recoded Categorical Variables into Binary Variables (continued)

Dwelling Characteristics	
roof1 =	1, Mud/mud bricks/earth, bamboo, palm leaves/thatch 0, Otherwise
roof2 =	1, Wood, slate/asbestos, roofing tile 0, Otherwise
roof3 =	1, Metal sheet 0, Otherwise
roof4 =	1, Concrete/Other 0, Otherwise
lighting1 =	1, Electricity or Solar 0, Otherwise
lighting2 =	1, Kerosene or gas lamp 0, Otherwise
lighting3 =	1, Flashlight/torch 0, Otherwise
lighting4 =	1, Candle, firewood etc. 0, Otherwise
toilet1 =	1, No facility 0, Otherwise
toilet2 =	1, WC 0, Otherwise
toilet3 =	1, Pit latrine 0, Otherwise
toilet4 =	1, KVIP 0, Otherwise
toilet5 =	1, Bucket/pan, public toilet etc. 0, Otherwise

Table 5.1: Recoded Categorical Variables into Binary Variables (continued)

Dwelling Characteristics	
wall1 =	1, Mud bricks/earth, Landcrete 0, Otherwise
wall2 =	1, Wood, Metal sheet/slate/asbestos, Bamboo, Palm leaves/thatch 0, Otherwise
wall3 =	1, Stone, Burnt bricks, Cement blocks/concrete, Other 0, Otherwise
floor1 =	1, Earth/mud 0, Otherwise
floor2 =	1, Cement/concrete, stone, burnt bricks 0, Otherwise
floor3 =	1, Wood, vinyl tiles, ceramic/porcelain tiles 0, Otherwise