MODELING MULTIDIMENSIONAL INDICATORS OF POVERTY IN AFRICA

George Mbogo Wamuyu
I56/68503/2013

This thesis is submitted in partial fulfillment of the requirements for the Degree of Master of Science in Social Statistics in the School of Mathematics

School of Mathematics
University of Nairobi
2015

Declaration

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

George Mbogo

Signed:

_____

Date:

_____

This thesis has been submitted for examination with my approval as the university supervisor.

Dr. Timothy K. K. Kamanu

Signed:

_____

Date:

_____

Acknowledgement

I give my thanks to the almighty God for the strength that he has given me to do this project.

Special thanks go to Dr.Timothy for the sound guidance he has rendered in the accomplishment of this work. I also thank the Demographic Health Program for granting me access to their data . I also give my heart felt gratitude to my family for their moral support and prayers.

Dedication

I dedicate this thesis to my family;My Mum,Wife and Daughter

Abstract

Poverty is a condition in which a person or community is lacking the basic needs required for a minimum standard of well-being. Poverty can be measured using economic, social and environmental indicators. Past research on measuring poverty has been done using an income approach. The income approach considers only one dimension of poverty and fails to consider other dimensions such as social and environmental dimensions. The aim of this study is to develop a concept of measuring poverty using a multidimensional approach.We used data from the Ethiopia Demographic Health Survey. Multi-Correspondence Analysis and Non Linear Principal Component Analysis were used to extract and weight indicators of poverty.The weighted indicators were then used to construct a Poverty index.This index can assist policy makers in quantifying poverty.

Contents

List of Tables

List of Figures

T H E S I S

Chapter 1

Introduction

Poverty is a condition in which a person or community is lacking the basic needs required for a minimum standard of well-being. Poverty is a plague afflicting people all over the world. Poverty is a manifestation of human deprivation and is linked to human capital underdevelopment. Poverty includes inadequate income and denial of the basic necessities such as education, health services, clean water and sanitation which are essential for human survival and dignity (World Bank, 2007).

Poverty can be measured using economic, social and environmental indicators. Measures of poverty focus on material needs, such as food, clothing, shelter, or safe drinking water.

Environmental measures of poverty include access to proper sanitation and water,access to proper living conditions, types of fuel used and access to good and productive land especially for rural populations. Social measures of poverty include lack of access to information, education, health care, or political power. Poverty is an aspect of inequitable social relationships, experienced as social exclusion, dependency, and/or diminished capacity to participate in society. Economic measures of poverty can be absolute or relative. Relative poverty is measured as the percentage of the population

with income less than some fixed proportion of median income.Absolute poverty refers to a set standard which is consistent over time and between countries.

The World Bank defines poverty in absolute terms.The bank defines extreme poverty as living on less than 1.25 US dollars per day, and moderate poverty as less than 2 US dollars a day. In order to compare international poverty lines, local poverty lines are converted to a common currency to give the purchasing power parity (PPP).This conversion is such that, one measures goods that can be purchased in the local currency equivalent to 1.25 US dollars for extreme poverty and 2 Us dollars for moderate poverty (Ravallion,1996).

A person is poor when their personal income or consumption is below a specified poverty line (Coudouel and Hentschel, 2000). However, personal income can vary greatly from year to year.The income approach is only appropriate for wage earners, and has less relevance to the poor. Many poor people rely on their own production and informal-sector activities in which the concept of profit is unclear, rather than on a formal income (Glewwe and Van der Gaag,1988).

1.1 Research Problem

1.1.1 Background of the Problem

Past research on poverty was conducted based on income measures. The rationale being that income is funds which can be spent to satisfy and fulfill basic human needs (Scott, 2002).

Recent research on poverty on the other hand is based on a more comprehensive concept, rather than the income-based approach. Poverty is now perceived and defined

based on financial, social and environmental indicators. Direct indicators enhance indirect indicators of poverty to obtain a comprehensive assessment of living conditions (Manuela, Mariangela, 2011).

1.1.2 Statement of the problem

Income measures of poverty have probably been used in the past because they are easy to compute and compare. However, income is not a sufficient measure. This study seeks to develop a measure of poverty that includes financial, social and environmental factors to define an alternative multiple dimensional index. The following concepts are emphasized.

- The concept of identification of relevant indicators of poverty
- The aggregation and weighting of the indicators of poverty.

1.2 Significance of the study

Development of a concept that can be used to identify indicators of poverty and a technique of aggregating these indicators to form an index is useful in the following ways.

- It will assist policy makers quantify poverty.
- it will help compare poverty levels.

The rest of this study is organized as follows. Chapter 2 discusses various approaches and social indicators used to measure poverty. The model used to weight and aggregate the indicators is discussed in Chapter 3. In Chapter 4, we discuss the results and Chapter 5 concludes the study.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

Income based measures of poverty are one dimensional i.e. they are based on a single indicator, income or expenditure per capita. Income based measures differentiate between poor and non-poor individuals on the basis of a poverty line which may be absolute or relative. Absolute measures are defined based on the amount of money one requires to earn a living (Nolan ,Whelan, 1996). Relative measures assess poverty from a threshold set at a specific percentage of median or mean income.Individuals based on the set specific threshold are unlikely to fully participate in life.

2.2 Absolute Approach

The problem of measuring poverty involves two exercises: the identification and aggregation of the characteristics of the poor into an overall indicator. The use of income method requires the specification of a subsistence income level referred to as poverty line. If a person falls below the line one is considered poor and the reverse is true. Sen proposed two measures of aggregation, the head count ratio (proportion of persons with income less than poverty line) and the income gap ratio (the gap between the poverty line and the average income of the poor expressed as a proportion of the

poverty line). Sen, noted that the income approach was insensitive to the redistribution of income among the poor. He suggested a more sophisticated index of poverty using an axiomatic approach (Sen,1976).

In a study carried out by Ravallion in 1990 using the absolute measure, the lowest mean consumption amongst the 86 countries studied in the World Development Report, was Somalia at twenty two US Dollars per person per month in 1985 PPP prices. A more generous, and more representative, absolute poverty line for low-income countries was thirty one US Dollars, which (to the nearest dollar) was shared by six of the countries in his sample.The countries were Indonesia, Bangladesh, Nepal, Kenya, Tanzania, and Morocco. In conclusion Ravallion concluded that a strong case could be made for treating a consumption level of twenty three US Dollars per person per month, in 1985 U.S.Dollar purchasing power parity as a reasonable lower bound for the poverty line. However, for comparative purposes it was worth considering a rather less meager criterion of a consumption level of thirty one US Dollars per month which was actually a far more common poverty line amongst the dozen or so low income countries for which poverty lines was calculated. The range US Dollar 23-31 was embraced quite well as poverty lines used by low income countries (Ravallion, 1990).

There are a number of contentious issues about the absolute approach; Income approach. The contention has been which the poverty indicators between consumption and income, should be included and how it should be valued. Another contention of the poverty line,is how it should vary between subgroups or dates and what level should be set on average. The are also questions on the poverty line on whether it should be additive, whether it needs to penalize inequality amongst the poor and how the resulting measure related to social welfare functions (Ravallion, 1996).

It can be agreed that even the best income and non-income measures found in practice are incomplete on their own. Considerable research has gone into the problem of identifying money metric utility from demand behavior, including setting equivalence scales which give the differences in income needed to compensate families with different demographic compositions (Ravallion,1996).

Ravallion further argued that by taking a multiple indicator approach there's no need, for each indicator to measure everything. It should be clear what exactly each is measuring, and why we need it. Four sets of indicators can be defended as ingredients. for a sensible approach to measuring poverty. These are

1. A sensible poverty measure based on the distribution of real expenditure per single adult covering all market goods and services including those obtained from non-market sources.

2. Non-Income indicators as access to non-market goods for which meaningful prices cannot be assigned, such as access to non-market education and health services.

3. Indicators of distribution within households; measures of gender disparities and child nutritional status.

4. Indicators of personal characteristics which entail unusual constraints on the ability of escape poverty, such as physical handicaps or impairments due to past chronic under nutrition.

A genuine measure of poverty should depend on income indicators as well as non income indicators that may help in identifying aspects of welfare not captured by incomes (Ravallion, 1996).

Indices should point out that command over market goods is not all that matters to

people's well being, other factors need to be considered when quantifying the extent of poverty and in informing policy making for fighting poverty. While assessing these indices a credible set of multiple indicators should be developed.The dimensions of poverty most relevant to a specific setting should be considered.When weights are needed, they shouldn't be set solely by an analyst measuring poverty.They should be consistent with well-informed choices made by poor people (Ravallion, 2011).

Income-based approaches or one dimension approaches to measurement of poverty are favored because of ease of use and computation. Nonetheless, the income based approach has been subjected to various criticisms of failing to encompass all dimensions of poverty. It assumes all persons are of earning age. This has led to attempts to find multidimensional indicators which can capture the different facets of poverty and deprivation. It has been increasingly recognized that other aspects of human life which are not necessarily related to income do impact on human development. These include access to public goods, health, education, housing conditions life satisfaction among others (Townsend, 1979).

The consequence of this evolution in the measurement of poverty has broadened the notion of poverty to include vulnerability, exposure to risks, voicelessness and powerlessness (World Bank, 2001). The definition of poverty is not limited to the lack of the ability of individuals/households to obtain sufficient resources to satisfy their basic needs neither is it considered on a mere economic and monetary dimension (Townsend, 1993). It is rather increasingly considered as human deprivation in various life domains. This deprivation from the multidimensional perspective includes both quantitative and qualitative measures such as the joy of choices, opportunities and others which are most basic to human development and can paint quite different pictures of the poverty situation in any given country (Alkire, 2002).

2.3 Multidimensional measures of poverty.

In the past few decades there has been a tremendous search for suitable approaches of measuring multidimensional poverty.

2.3.1 Union or Intersection Approach

Atkinson, introduced a concept of measuring poverty which is known as the intersection or union approach.The union approach assumes one who is deprived in a single dimension, is regarded as poor in a multidimensional setting. However, this is overly inclusive and may lead to exaggerated estimates of poverty. In the intersection approach, one is required to be deprived in all dimensions before being identified as poor. This approach too is often to constricting and usually produces very low estimates of poverty.

2.3.2 Human Development Index (HDI)

Another concept of measuring poverty is the Human Development Index (HDI). The HDI is a composite index measuring deprivations in the three basic dimensions. a long and health life, knowledge and a decent standard of living (Haq,1995).

The calculation of HDI involves three dimensions health (h), education (e), and the ability to achieve a decent standard of living, represented by income (y). The performances of each country in these three dimensions are normalized, then aggregated to get the composite HDI. Prior to 2010, linear averaging (LA) across three dimensions was used as an aggregation method to obtain HDI. In 2010, this aggregation method was revised to the geometric mean (GM) (UNDP, 2010).

This index has been criticized for attributing arbitrary equal weights to each dimension. Secondly,the choice of which variables should be included in the HDI is somehow arbitrary and may not reflect peoples' preferences and realities in the country under study (Booysen, 2002).

2.3.3 Dual Cut off method

Alkire and Santos proposed a methodology of computing the multidimensional poverty index (MPI), where poor households are identified and an aggregate measure. The MPI has three dimensions: health, education, and standard of living. These are measured using ten indicators. Each dimension is equally weighted; each indicator within a dimension is also equally weighted. (Alkire, Santos 2010a),
The MPI reveals the combination of deprivations that batter a household at the same time. A household is identified as multi-dimensionally poor if, and only if, it is deprived in some combination of indicators whose weighted sum is 30 percent or more of the dimensions. The dimensions, indicators, and deprivation criteria are presented below and explained with below.

1. Health (each indicator weighted equally at$\frac{1}{6}$).
    (a) Child Mortality: If any child has died in the family.
    (b) Nutrition: If any adult or child in the family is malnourished.
2. Education (each indicator weighted equally at$\frac{1}{6}$).
    (a) Years of Schooling If no household member has completed 5 years of schooling.
    (b) Child School Attendance If any school-aged child is out of school in years 1 to 8.
3. Standard of living (each of the six indicators weighted equally at$\frac{1}{18}$).

(a) Electricity If household does not have electricity.

(b) Drinking water If does not meet MDG definitions, or is more than 30 mins walk.

(c) Sanitation If does not meet MDG definitions, or the toilet is shared.

(d) Flooring If the floor is dirt, sand, or dung.

(e) Cooking Fuel If they cook with wood, charcoal, or dung.

(f) Assets If one does not own more than one of: radio, tv, telephone, bike, motorbike or refrigerator and do not own a car or truck.

The MPI is the product of two numbers: the Headcount H or percentage of people who are poor, and the Average Intensity of deprivation A which reflects the proportion of dimensions in which households are deprived. Alkire and Foster show that this measure is very easy to calculate and interpret, is intuitive yet robust, and satisfies many desirable properties.

2.3.4 Multivariate Techniques

Multivariate techniques can be used in the measurement of poverty.In particular,the Principal Component Analysis method of optimal scaling can be used to set weights used in the poverty index construction.

The Principal Component Analysis (PCA) reduces data dimension. It consists building a sequence of uncorrelated (orthogonal) and normalized linear combinations of input variables (K primary indicators).These uncorrelated linear combinations are latent variables called "components". The optimality of the PCA process is such that the 1st component has maximal variance, and all subsequent components have decreasing variances whose sum is the total variance of the K indicators. This total variance is also named the total inertia of the distribution of the K indicators.

The PCA technique has some limitations. It has been developed for a set of quantitative variables, measured in the same units. The optimal sampling properties for parameter estimation depend on the multivariate normal distribution and do not exist with qualitative variables.

2.4 Social Indicators

Social indicators are based on observations. These are measures that describe the well being of individuals or communities. They are used to describe and evaluate community well-being in terms of social, economic and environmental welfare.

2.4.1 Age and Sex

The age group and sex classification of social indicators are used to detail various characteristics of individuals. Infants, children youth and the elderly are mostly dependent persons. The elderly aged 60 and above are considered in to categories. The group is divided in to two groups aged 60-69 to represent a relatively active and self sufficient period and 70 and over, when health disability, income and social issues are likely to become more pronounced. Persons in the age of 25 to 59 are considered to be productive persons who can meet their needs. An age break at 45 is useful for distinguishing the reproductive and post productive ages of women as well as general periods of the adult life cycle characterized by relatively different patterns and rate of change in labour force, household and marital characteristics and child dependency. (UN, 1989)

2.4.2 Urban and Rural

The distinction between urban and rural is mainly based on the size of locality. The basic assumption is that urban areas provide a different way of life than is found in the rural areas. This classification is useful and significant for initial and essen-

tial disaggregation of national data for indicators where greater disaggregation is not feasible. The basic importance of the urban and rural classification derives from two considerations. First, it is the only geographical classification at the international level and in many cases at the national level which is readily available and which can be used to identify with country social and economic differences .Secondly it's the most feasible way of obtaining data on the rural population which is of fundamental importance for national policies and programmes concerned with agrarian reform and rural development. (UN, 1989)

2.4.3 House hold size and composition; Household leadership

A household is based on the arrangement made by persons, individually or in groups, for providing themselves with food or other essentials for living. A household may be either a one person household, that is, a person who makes provision for his or her own food or other essentials for living without combining with any other person to form a multi-person household. A multi-person household is a group of two or more persons living together who make common provision for food or other essentials for living. These person may pool incomes and have a common budget to a grater or lesser extent; they may be related or unrelated or a combination of both (UN, 1989). The concept of family is more restricted than the concept of household in that a household may consist of only one person but a family must contain at least two members, and the members of a multi-person household need not be related to each other while the members of a family must be related. In identifying members of a house hold it is traditional first to identify the household or family head then the remaining of the household or family according to their relationship to the head. The head of the household or the family is defined as that person in the household or the family who is acknowledged as such the other members. The head of household is identified based on the fact that most household are family house holds that is they consist entirely of persons related by blood, marriage, or adoption save for domestic servants. The assumption is that the head of the household has primary authority and

responsibility for household affairs and, in the majority is its chief economic support. This classification is useful to determine the household size, households with members under age 15 and sex of the head. It's also useful in determining other household characteristics such as age of head, number of family nuclei present, presence of non-related persons and a number of generations in the household.

2.4.4 Level of Education

Education, besides being a source of potential enjoyment in its own right, confers many advantages and tends to affect attitudes and points of view of an individual. .It is assessed in several ways e.g. the number of years in the education system, the types of institutions attended, subjects studied and qualifications obtained (UN, 1975). Education attainment refer primarily to the highest grade completed within the most advanced level attended in the educational system of the county where the education was received, but it should also take into account any adult education measurable in levels and grades or equivalent even if it was provided outside of the regular school or university program. For international purposes, a grade is a stage of instruction usually covered in the course of a school year (UN, 1989). The UNESCO classification of education by level as it is applied in the United Nations population census recommendations is given below:

1. Education at the first level, which usually begins between ages 4 and 7 and lasts about 5 years

2. Education at the second level, which begins at about age 10 -12 and lasts for about 3 years

3. Education at the second stage of the second level, which begins at the about age 13-15 and lasts for about 4 years;

4. Education at the third level, which begins at about 17-19 and lasts for at least 3 or 4 Years or longer, depending upon the stage. This level entails college and university education.

Occupation and Status in employment

Occupation refers to the kind of work done during a selected time(reference period) or kind of work done in the past.If employed, irrespective of the industry in which an individual works or his or her status. The status in employment refers to the status of an economically active individual with respect to his or her employment. The international recommended status in employment classification is contained in the United Nations population census recommendation, and is fundamental for distinguishing at least approximately relatively organized economic activity(employers and employees) from small-scale household economic activity(unpaid family workers and own account workers) in the various branches (UN,1989).

According to this classification, an employer is a person which operates his or her own unincorporated economic enterprise or engages independently in a profession or trader and hires one or more employees. An own account worker is a person who operates his or her own economic enterprises or engages independently in a profession or trade and hires no employees. An employee is a person who works for a public or private employer and receives remuneration in wages, salary, commission, tips and piece-rates or pay in kind. An unpaid family worker is a person who works without pay in an economic enterprise operated by a related person living in the same household.

2.4.5 Time use and Economic Activities.

Time use is an indicator to provide data on economic activity, household work and other activities, distinguishing agricultural, market and non market activities and, in the field of leisure and free time education and social activities (UN, 1989). The kind of economic activity classification refers to the principal type of economic activity in which an economic production unit is engaged, whether the unit is a public or private enterprise or establishment, a household, and individual working on own account or unit of government.

Economic indicators are classified in to the following categories;

- Agriculture/agricultural services and forestry/fishing are distinguished because of their great importance in developing countries and the difference economic and social arrangements associated with each group of activities.
- Trade and economic services are distinguished for the same reasons particularly to elucidate the role of petty trade in the so called informal sector;
- Recreational and cultural services are an important group for indicators in the field of leisure and culture;
- Personal services are also important in many developing countries.

2.4.6 Housing

Housing, dwellings or living quarters consists of all separate and independent premises including vacant premises used for human habitation whether or not they were originally designed for that purpose (UN, 1975). Housing units are of many kinds ranging from palaces to caves, they also entail living quarters which are not housing units such as hotels, rooming houses and other lodging houses, camps and institutions.
Housing is evaluated in the following classifications:

- Durability of housing: A house is considered ,durable. If it is built on a non-hazardous location and has a structure permanent and adequate enough to protect its inhabitants from the extremes of climatic conditions, such as rain, heat, cold and humidity.
- Sufficient living area: A house is considered to provide a sufficient living area for the household members if not more than three people share the same habitable (minimum of four square meters) room.

2.4.7 Water: Access to improved water

Improved drinking water technologies are more likely to provide safe drinking water than those characterized as unimproved. A household is considered to have access to an improved water supply if it uses improved drinking water sources or delivery points (listed below).

Improved drinking water sources include:

Piped water into dwelling, plot or yard; public tap/standpipe; tube well/borehole; protected dug well; protected spring; wainwater collection.

Unimproved drinking water sources include:

Unprotected dug well, unprotected spring, cart with small tank/drum, bottled water; tanker-truck, surface water (river, dam, lake, pond, stream, canal, irrigation channels).

2.4.8 Sanitation: Access to improved sanitation

Improved sanitation facilities are more likely to prevent human contact with human excreta than unimproved facilities. A household is considered to have access to improved sanitation if it uses improved sanitation facilities (listed below).

Improved sanitation facilities include:

Flush or pour-flush to piped sewer system, septic tank or pit latrine, ventilated improved pit latrine, pit latrine with slab; and composting toilet.

Unimproved sanitation facilities include:

Flush or pour-flush to elsewhere, pit latrine without slab or open pit, Bucket, hanging toilet or hanging latrine, no facilities or bush or field.

2.4.9 Cooking Fuel

Bio Mass Fuel refers to burned plant or animal material; wood, charcoal, dung and crop residues. The polluting effect, efficiency and cost of domestic fuel use are often construed as an energy ladder(WHO, 2006a).

Dried animal dung, scavenged twigs and grass, which are cheap, inefficient pollute

the most, are at the bottom of the ladder. Crop residues, wood and charcoal have a higher level BMF, whilst kerosene, coal and bottled or piped gas are the most efficient (non-BMF) combustible energy sources. Electricity is at the top of the energy ladder. The correlation of socioeconomic factors with the main fuel used is relatively close, however most households use several fuels in different settings.

Four factors that appear to be most relevant in a household's choice of fuel type are: Cost of fuel, stove type and accessibility to fuels, technical characteristics of stoves and cooking practice, cultural preferences; and lastly, if at all, the potential health impacts (Masera et al., 2000).

Chapter 3

METHODS

This chapter describes the data, variables and models considered in the study.

3.1 Introduction and Data description

This study uses data collected in the Ethiopia Demographic and Health Survey (EDHS) at national and regional levels. The data comprises a sample of 17817 individuals. information relating to their demographic characteristics, information on asset ownership, and access to proper sanitation, water and housing characteristics was collected. The survey has been done using a questionnaire to collect the data.

3.2 Multi Correspondence Analysis.

Multiple correspondence analysis (MCA) is an extension of correspondence analysis (CA).It analyzes the pattern of relationships of several categorical dependent variables.Correspondence analysis is a technique for representing the information in a two-way contingency table, which contains the counts (frequencies) of items for a cross-classification of two categorical variables.We can construct a plot that shows the interaction of the two categorical variables along with the relationship of the rows to each other and of the columns to each other.

MCA is a carried out on an indicator (or design) matrix with cases as rows and categories of variables as columns. The inner product of the design matrix is Termed the Burt Table

MCA is used to analyze a set of observations described by a set of nominal variables. Each nominal variable comprises several levels, and each of these levels is coded as a binary variable. For example gender, (F vs M) is one nominal variable with two levels. The pattern for a male respondent will be 0 1 and 1 0 for a female. The complete data table is composed of binary columns with one and only one column taking the value 1 per nominal variable. MCA can also accommodate quantitative variables by recoding them as bins. For example, a score with a range of-5 to 5 could be recoded as a nominal variable with three levels: less than 0, equal to 0, or more than 0. With this schema, a value of 3 will be expressed by the pattern 0 0 1. The coding schema of MCA implies that each row has the same total, which for CA implies that each row has the same mass .

The goal of idea in CA is to reduce the dimensionality of a data matrix and visualize it in a subspace of low-dimensionality, commonly two- or three dimensional. The data of interest in CA are usually a two-way contingency table or any other table of nonnegative ratio-scale data for which relative values are of primary interest.

To summarize the theory, first divide the $I \times J$ data matrix, denoted by N, by its grand total n to obtain the so-called correspondence matrix $P = N/n$.Let the row and column marginal totals of P be the vectors r and c respectively, that is the vectors of row and column masses, and D_r and D_c be the diagonal matrices of these matrices.

The computational algorithm uses singular value decompostion to obtain coordinates of the row and column profiles with respect to principal axes, using the Singular Value Decomposition (SDV), is as follows:
Calculate the matrix of standardized residuals:

$$S = D_r^{-1/2}(P - rc^T)D_c^{-1/2}. \tag{3.1}$$

Calculate the SVD:

$$S = UDV^T \qquad \text{where} \qquad U^TU = V^TV = I. \tag{3.2}$$

Principal coordinates of rows:

$$F = D_r^{-1/2}UD_\alpha \tag{3.3}$$

Principal coordinates of columns

$$G = D_c^{-1/2}VD_\alpha \tag{3.4}$$

Standard coordinates of rows:

$$X = D_r^{-1/2}U \tag{3.5}$$

Standard coordinates of columns:

$$Y = D_c^{-1/2}V \tag{3.6}$$

The significance of the data matrix is measured by the inertia which is a chi-square

statistic. It is calculated on relative observed and expected frequencies:

$$\text{Inertia} \equiv \delta^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(p_{ij} - r_i c_j)}{2\, r_i c_j} \tag{3.7}$$

The rows of the coordinate matrices in (3.3)-(3.6) above refer to the rows or columns, of the original table, while the columns of these matrices refer to the principal axes, or dimensions, of the solution. The row and column principal coordinates are scaled in such a way that $FD_rF^T = GD_cG^T = D_\alpha^2$ i.e. the weighted sum-of-squares of the coordinates on the k-th dimension (inertia) is equal to the principal inertia (eigen value) α_k^2 ,*the square of the k-th singular value*, whereas the standard coordinates have weighted sum-of-squares equal to 1: $XD_rX^T = YD_cY^T = I$

3.3 Non Linear Principal Component Analysis

Non Linear Principal Component Analysis is the non linear equivalent of standard Principal Component Analysis.It reduces the observed variables to a number of uncorrelated principal components .It handles nominal and ordinal variables to discover nonlinear relationships between variables. NLPCA is able to perform analysis of likert type scales(variables with levels eg.Very good, Good, Bad Worse) using optimal quantification.

- If the dataset of interest is collected from n objects with m categorical variables. Each variable has kj categories which have certain measurement level possibly numerical, ordinal or nominal. The main interest here is to display these objects and variables in a joint p-dimensional space in a way that.
- Objects with similar profiles are close together.

- Categories with similar contents are close together.

We achieve maximum homogeneity by quantifying or re-scaling the objects and the variables. This depicts the leading idea on which the homogeneity analysis is based.

In order to quantify both the objects and the variables, the original dataset has to be coded in to an indicator matrix G. The rows and columns of the indicator matrix represent the n objects and m variables respectively. For j-th variables, Gj, a $n \times k_j$ indicator matrix, is assigned to denote those kj categories of the variable and if the s-th object belongs to t-th category then the entry gs,t = 1, otherwise gs,t = 0.After integrating all G_j , we have $G = [G_1, G_2, G_3, G_m]$. We also assign Y_j, a k_j x p matrix, as the multiple category quantification of j-th variable and x, a $n \times p$ matrix, as the resulting optimal scores of objects.

The orthogonal projector is defined as $P_j = G_j D_j^{-1} G\prime j$. where $D_j = G\prime_j G_j$, and the definitions of perfection are as follows:

1. Yj (Category quantification)are perfectly homogeneous if;
   $G_1 1 Y_1 = G_2 Y_2 = . = G_m Y_m$
2. X (Object scores) are perfectly discriminated if
   $X = P_1 X = P_2 X = \ = P_m X.$
3. X and Yj are perfectly consistent if
   $X = G_1 Y_1 = G_2 Y_2 = .. = G_m Y_m$

The NLPCA optimal quantification task and the linear PCA model estimation are performed simultaneously, which is achieved by the minimization of a least-squares loss function.

It is rare to find the solutions for both X and Yj that achieve perfect homogeneity,instead we find the solutions which minimize the departures from perfect homogeneity. To measure departures of perfect homogeneity, we define the loss function:

$$\delta^2(X; Y_1, \ldots, Y_m) = \frac{1}{m} \sum_{j=1}^{m} tr(X - G_j Y_j)'(X - G_j Y_j) \tag{3.8}$$

The minimization problem is solved by using the iterative alternating least squares algorithm (ALS) also referred to as the reciprocal averaging algorithm Leeuw (2009).

At iteration t =0, we start with an arbitrary object scores$X(0)$. Each iteration t consists of three steps:

Update category quantifications:

$$Y_j^t = D_J^{-1} G_j' X^t \quad \text{for} \quad j = 1, \ldots, m \tag{3.9}$$

Update object scores:

$$X^t = M_*^{-1} \sum_{j=1}^{J} G_j Y_j^t \tag{3.10}$$

Normalization:

$$X^{t+1} = M_*^{-1/2} orth(M_*^{-1/2} X^t) \tag{3.11}$$

We repeat step 1, 2 and 3 until X and Y are converged.

Chapter 4

DATA ANALYSIS AND FINDINGS

4.1 Introduction

The data was obtained from the Demographic and Health Surveys Program in the SPSS format. The analysis of data has been done using R software.Several R packages have been used in the analysis. The R package Foreign has been used to import the data from SPSS into R. The Homals package has been used to implement Non-Linear Principal Component Analysis. The FactoRmine and ca packages have been used to implement Multi Correspondence Analysis.

In the analysis of data, variables that were relevant to this thesis from the Ethiopia Demographic Health Survey dataset were chosen.The variables were put in categories before starting the analysis. The variables put in categories were further edited by combining some of their groups in one or two groups either because of the small number of observations in those categories or to make the analysis and the interpretation more meaningful.

4.2 NLPCA Results

To interpret the data. it is necessary to reduce the number of variables to a few, interpretable linear combinations of the data. Each linear combination will correspond to a principal component.

To interpret each component, we must compute the correlations between the original data for each variable and each principal component.These correlations between the principal components and the original variables are used to interpret these principal components.Interpretation of the principal components is based on finding which variables are most strongly correlated with each component, i.e., which of these numbers are large in magnitude, the farthest from zero in either positive or negative direction.The criteria of deciding which numbers are considered to be large or small is of course a subjective decision.

Table of components and original variables

Scores	Component1	Component2
Toilet.facility	-0.11309	-0.00022
Main.Floor.Material	-0.14829	0.01138
Main.Wall.Material	-0.06936	0.01254
Main.Roof.Material	-0.14708	-0.00963
Cooking.Fuel	-0.11883	0.00808
Electricity	0.17023	0.01303
Radio	0.06638	-0.07932
TV	0.14165	-0.03248
Refigerator	0.06951	-0.04717
Bicycle	0.00781	-0.03256
Motorcycle.Scooter	0.00917	-0.02349
Car	0.01564	-0.02241
Relationship.Structure	-0.05338	-0.13798
Gender	-0.03303	-0.12483
Telephone	0.08435	-0.04420
Share.Toilet	0.09694	0.06410
Person.fetching.water	-0.08117	0.00156
Water.safety	-0.03874	0.02711
Separate.Kitchen	-0.00266	-0.04304
Mobile.telephone	0.13008	-0.03847
Watch.clock	0.03424	-0.09818
Animal.cart	-0.00186	-0.01523
Agricultural.land	-0.14250	-0.03341
Own.livestock	-0.13777	-0.05275
Bank.Account	0.07727	-0.04729

From the results of the NLPCA Considering the 1st Principal component, Toilet facility, Main Floor Material, The roof Material, electricity, Television, mobile

phone, agricultural land and livestock are more strongly correlated to the Principal component.

The second principal component is also more correlated to the relationship structure and gender. The criteria of selecting the principal components is any variable with a coefficient of 0.1 and above.

4.3 MCA Results.

The first principal component accounts for 57.8 percent of the variation. In the construction of the index we share use the factor scored that are correlated with the first principal component.

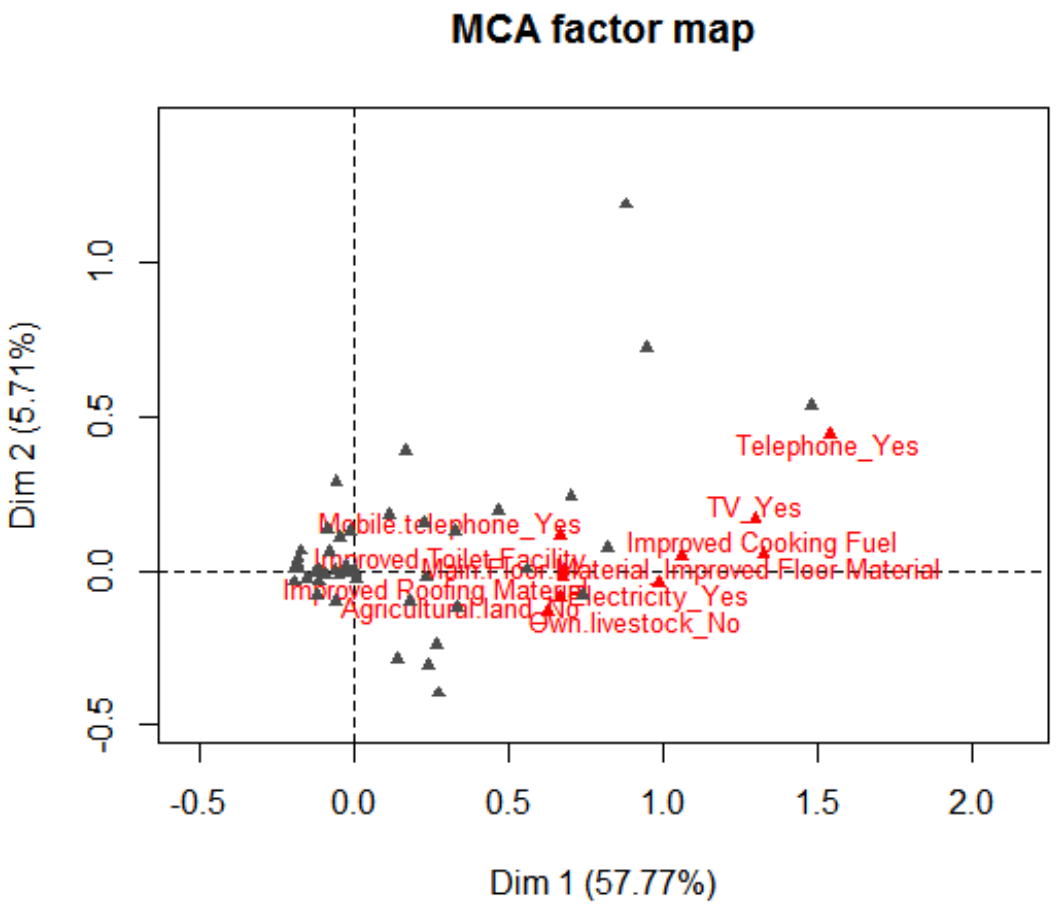
Table 4.1: Table of components and the variability they account for

Dimension	value	Percentage	Cum Percentage
1	0.05505	57.8	57.8
2	0.00544	5.7	63.5
3	0.00253	2.7	66.1
4	0.0023	2.4	68.6
5	0.00207	2.2	70.7
6	0.00188	2	72.7
9	0.00158	1.7	78
10	0.00154	1.6	79.7
11	0.00152	1.6	81.3
34	6.90E-50	0.1	100

The MCA factor plot below also illustrated the factors that contributed the most variation in the model.

In the analysis of a factor plot, Variable that are close in the plot are deemed to be more correlated.In the factor plot above, Toilet facility, Main Floor Material, The roof Material, electricity, Television, mobile phone, agricultural land

Figure 4.1: Multiple Correspondence Analysis Factor map



and livestock are identified as having a strong correlation.In the above plot variables with lower correlations were left out and plotted in colour grey to enhance visualization.

In the table below MCA, was used to identify the variables which contribute the highest inertia.The ten variables in the table that gave the most variation from our data.

Table 4.2: Table of scores and original variables

Categorical variables	Dimension 1	Dimension 2	Dimension 3
Drinking.Water	0.19	0.006	0.033
Toilet.facility	0.333	0.000	0.007
Main.Floor.Material	0.572	0.003	0.001
Main.Wall.Material	0.125	0.003	0.075
Main.Roof.Material	0.562	0.002	0.007
Cooking.Fuel	0.367	0.002	0.002
Electricity	0.753	0.004	0.01
Radio	0.115	0.169	0.021
TV	0.522	0.027	0.011
Refigerator	0.126	0.052	0.233

4.4 Calculating the Poverty Index

In order to reduce the dimension, both NLPCA and MCA we have been used to identify the variables that contribute the most variability.These variables are used to construct the poverty index.

The Poverty index has been constructed using the weights obtained from the results of MCA.The initial step in computing the poverty index was to estimate the scores.We then use category variables identified as contributing most variability using MCA method. The scores are multiplied by the number of responses for each variable.

The following equation has been used to calculate the Poverty index score for each population unit

$$MCA_{Pi} = R_{i1}W_1 + R_{i2}W_2 + \cdots + R_{ij}W_j + \cdots + R_{iJ}W_J \tag{4.1}$$

Where MCA_{Pi} is the ith households composite poverty indicator score,R_{ij} is

the response of Household i to category j, and Wj is the MCA weight for dimension one applied to categoryj. The resulting index was the categorized as follows.Very Poor 0-0.2, Poor 0.3-0.4, Middle Income 0.5-0.6,Rich 0.7-0.8 and very rich 0.9-1.0.

From Table 4.3 The poverty index for Ethiopia is 0.33. We conclude that Ethiopia is a poor country.

Table 4.3: Calculation of the Poverty Index

Variables	**Categories**	**Weights**	**Response**	**RiWi**
Water Source	Improved	0.482	1350	650.7
	Unimproved	-0.394	1650	-650.1
Toilet	Improved	1.393	439	611.527
	Unimproved	-0.239	2561	-612.079
Main Floor Material	Improved	2.184	321	701.064
	Unimproved	-0.262	2679	-701.898
Main Wall Material	Improved	1.689	126	212.814
	Unimproved	-0.074	2874	-212.676
Roofing Material	Improved	1.392	675	939.6
	Unimproved	-0.404	2325	-939.3
Cooking Fuel	Improved	2.728	141	384.648
	Unimproved	-0.135	2859	-385.965
Electricity	No	-0.371	2537	-941.227
	Yes	2.032	463	940.816
Radio	No	-0.25	1941	-485.25
	Yes	0.458	1059	485.022
TV	No	-0.195	2796	-545.22
	Yes	2.674	204	545.496
Refrigerator	No	-0.041	2960	-121.36
	Yes	3.049	40	121.96
Telephone	No	-0.058	2946	-170.868
	Yes	3.177	54	171.558
Share Toilet	No	-0.357	1972	-704.004
	Yes	0.685	1028	704.18
Mobille.Telephone	No	-0.319	2435	-776.765
	Yes	1.377	565	778.005
Watch/Clock	No	-0.131	1914	-250.734
	Yes	0.232	1086	251.952
Animal Cart	No	0.001	2982	2.982
	Yes	-0.122	18	-2.196
Agricultural Land	No	1.375	655	900.625
	Yes	-0.384	2345	-900.48
Owns Livestock	No	1.291	685	884.335
	Yes	-0.382	2315	-884.33
Owns a Bank account	No	-0.107	2793	-298.851
	Yes	1.447	207	299.529
Total RiWi				994.06
N				3000
PI				0.33

Chapter 5

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

Multi Correspondence Analysis (MCA is a good method of identifying and Weighting the indicators of poverty and in the construction of an index that can measure poverty. This Method of Analysis is useful in assessing poverty using indicators such as asset ownership, access to information, type of housing, access to safe water and sanitation etc. These indicators can be used to assess the poverty status.MCA, is ideal to model categorical data or discrete data.

5.2 Recommendations

More effort is required to develop measures of poverty to inform poverty makers accordingly.

Bibliography

[1] Atkinson,A.(2003) *M*ultidimensional Deprivation: Contrasting Social Welfare and Counting Approaches.Journal of Economic Inequality, 1(1): 5165.

[2] Ravallion M (1996). *I*ssues in Measuring and Modelling Poverty.The economic Journal 106,1328-1348.

[3] Leeuw J, Mair P (2009). *G*ifi Methods for Optimal Scaling in R.The Package Homals.The Journal of Statistical Software 31(4)

[4] Sebastien L, Julie J, Francois H (2008). *A*n R Package for Multivariate Analysis.The Package Factormine. *The Journal of Statistical Software*, 25(1).

[5] Nenadic O, Greenacre M. (2007). *C*orrespondence Analysis in R, with Two and Three dimensional Graphics. *The journal of Statistical Software*, 20(3).

[6] World Bank (2007). *Poverty at a glance.* World Bank, Washington, DC.

[7] Manuela and Mariangela (2011). *D*eriving Multidimensional Poverty Indicators: Methodological Issues and an Empirical Analysis for Italy.Social Indicators Reserach 107(1), 37-54

[8] Scott, L. (2002). *A* Poverty Indicator System for Local Government.Development Southern Africa 19, 483501.

[9] Vijaya K. (2010) *Constructing an area based socio economic status index.A principal component analysis approach.*Paper presented at the Early Childhood Intervention (ECIA) 2010 Conference

[10] Booysen F, Van Der Berg S et al (2010). *Using an asset index to assess trends in poverty in seven Sub-Saharan African Countries.* Paper presented at the Multidimensional Poverty conference.29-31 Agust Brasilia, Brazil

[11] Booysen, F. le R. (2002). *Poverty and Health in Southern Africa.*University of Stellenbosch

[12] Townsend, P. (1993). *The International Analysis of Poverty.*Routledge New York

[13] Nussbaum, Martha; Sen. Amartya(1993). *The quality of life.*Oxford England New York:Clarendon Press Oxford University Press

[14] Sen. A.(1976). *P*overty, an ordinal approach to measurement.Econometrica 44(2): 219-231

[15] UNICEF (2010). *State of the Worlds Children.*UNICEF p.18-19.

[16] UNICEF (1999). *The State of the Worlds Children.* UNICEF.

[17] UNDP(2007). *Human Development Report (HDR).* UNDP p.25.

[18] UN(2006). *United Nations Human Development Report.* UNDP p.6, 7.

[19] UN(2005). *State of the Worlds Children.* UNICEF.

[20] UN Millenium Project (2005). *Investing in Development. A Practical Plan to Achieve the Millennium Development Goals.*New York.

[21] World Bank(2000). *World Development Report 2000/01:Attacking Poverty.*New York University Press.