SMALL AREA ESTIMATION with Application to child malnutrition in Rwanda

Ildephonse NIZEYIMANA
Reg.Number: 214003403

College of Science and Technology

School of Pure and Applied Science

Master of Science in Applied Mathematics

Huye; November 14, 2016

SMALL AREA ESTIMATION with Application to child malnutrition in Rwanda

by

Ildephonse NIZEYIMANA
Reg.Number: 214003403

A dissertation submitted in Partial fulfillment of the requirements for the degree of
Master of Science in Applied Mathematics-Statistical Modelling and Actuarial Sciences

In the College of Science and Technology

Supervisor: Innocent NGARUYE

Huye; November 14, 2016

Declaration

This is to certify that this thesis is my own work. It has not been presented elsewhere for an academic reward. The references used are mentioned as recommended.

Date : November 14, 2016

Student's signature:

Student's name: Ildephonse NIZEYIMANA

Dedication

To:

The Most High;
My mother and my sisters;
My friends and colleagues;

I dedicate this work.

Acknowledgment

I thank The Almighty God for His daily protection upon me and for having enabled me to accomplish this work. I thank my supervisor Innocent NGARUYE for his guidance in this work. His contribution is appreciated because he was overloaded but he patiently continued helping me. Because of having got much more knowledge about statistics, I thank both my supervisor and the examiner Martin Singull to have suggested this topic of Small Area Estimation in which I have been very interested and leave me with more knowledge.
I thank Dr. MINANI Froduald, the coordinator of Masters program for the encouragement he gave us when sharing views. He helped me to be more determined. In general, I thank the whole team of lecturers who taught us during this program and the University of Rwanda to have given us some facilities in our studies when it was needed and possible.
I wholeheartedly thank my mother, NGIRIRABATANYURWA Immaculate, to have endured all sufferings and sacrificed herself so that I may become who I am now. She has done what exceeded her ability and deprived herself some life necessities for my welfare and future. Her education from childhood with advices of wisdom helped me to have life guidelines and shape.
I thank my colleagues with whom we shared life experiences and who contributed to improve my behavior towards the society. I also acknowledge the encouragement of my different friends in this journey of studies.
May The Almighty God grant all of them blessings from Above!

Abstract

In this thesis, we were aiming at investigating the small area estimates of the malnutrition indicators. For this, we first studied the small area estimation technique using Elbers, Lanjouw and Lanjouw (ELL) method. This approach is used in the case of lack of precision and accuracy of direct estimates at a sub-population level. The sub-population level meant here can be a geographical region or a demographical domain. To solve this problem, we introduce other variables called covariates to borrow strength in other related (mainly neighboring) domains or past researches in the same domain. Such variables must be such that they usually are in correlation with the variable of interest. In our study, we considered the districts to be the small areas. The malnutrition indicators which were our interest are: Stunting also called height-for-age, Underweight also called weight-for-age and Wasting also called weight-for-height. The covariates used to explain the rates of those indicators are: poverty rate, illiteracy rate and urbanization rate. These three covariates were chosen because they were found to be of great impact on nutritional status as published in the report of National Institute of Statistics Of Rwanda (NISR) of the 2010 Demographic and Health Survey. Some information was obtained from the third integrated household living conditions (EICV3) report. Stunting which is the indicator of chronic malnutrition was found to be of the highest proportion compared to other indicators in all districts. The underweight indicating the malnutrition in the recent period of time comes at the second place and lastly, the wasting. Apart from the reduction of malnutrition indicators rates which may have undergone in general, the same remarks are highlighted. The districts from Kigali City province are not much suffering from malnutrition followed by those of the Eastern province. Among six first suffering districts referring to stunting, there are three of the Western Province, two of the Southern Province and One of the Northern Province. Those districts are: Nyamasheke, Rutsiro, Ngororero, Gisagara, Nyaruguru and Burera. As stunting was at low proportion in Kigali City, the underweight is also more than two times lower than for all other Provinces. The wasting proportion is not high in general in the whole country.

Contents

1. Introduction

1.1. General Introduction

In recent years, Rwanda has made an significant progress in different domains including health domain among others. Rwanda is a small and landlocked country located in the Central and Eastern Region of Africa with surface area of 26,338km^2. Rwanda is one among countries which are developing faster even if it is still among low-income, food-deficit and least-developed countries.
In developing countries, food insecurity is a major problem and consequently the malnutrition which is manifested especially in children and in child-bearing age and lactating mothers.
Since the life of a child begins from his/her conception, that is why many efforts were made for the welfare of mothers. Good health of Child-bearing age mothers and especially children under five years old is a major concern of the Government of Rwanda.

Child-bearing age mothers and children under 5 years old are the most vulnerable and are taken into special consideration. Life conditions of a child under five years old must be improved because this is the period when the human body builds its immunity. In addition, a malnourished child under five years will tend to have problems in his growth and even low mental ability. Life conditions of a pregnant mother are also reflected to the child to be born.

In addition to the good will of the Government of Rwanda (GOR), some Non-governmental organizations (NGOs) are committed to eradicate malnutrition of children. However, the resources are always a major problem to achieve this mission. It is common that in many (if not all) projects there might be the budget constraint. However, there is not only the problem of insufficiency of food, but also its utilization (when available) was found to be a major problem in many families.

Negative effects of malnutrition are observed in school performance because; according to the study conducted in 2013 [2]; about 327,500 children who were found to have repeated a class, 13%(44,255 students) of this number was associated with stunting. In 2012, Rwanda was expected to have a Gross Domestic Product (GDP) loss because of short productivity of the adult working age population estimated at 49% of the whole population having had suffered from chronic malnutrition. Consequently; Rwanda loses from malnutrition 11.5% of her GDP in general [2].

Rwandas Vision 2020 strategy, with its focus on good governance, productive and market-oriented agriculture, and regional and international economic integration, has been a key to development. Economic growth has been strong, with GDP growth averaging of 8 percent over the past decade where it was 8.8% in 2011. The poverty rate dropped from 77.8 % in 1995 to 56.7% in 2006 and then to 44.9 % in 2012 [7]. An important remark is that the rural poverty rate is three times higher than that in urban populations.

In 2010, there were improvements in nutritional status among children under 5 and among women compared to the previous five years. Community health workers were put into place to help people. Rates of stunting, underweight and wasting have all decreased, and there has been a remarkable reduction in anemia in children under 5. Further, in September 2013, the Government of Rwanda (GOR) together with development partners launched a 1,000 days campaign that would be implemented in three phases to increase awareness of improved maternal, infant and young child feeding practices [3],[4].

Despite of remarkable efforts made for the eradication of malnutrition, the situation of child stunting in 2013 was found still serious with the highest rates (58%) among children 6-18 months of age. Almost 15 % were found to be stunted at two months which indicates a poor growth of the foetus during pregnancy. Underweight prevalence for children under five years of age in Rwanda was 3.6 % nationally in 2012. The prevalence was 12% for children 6-12 months.
In general, the prevalence of stunting prevalence among children under five years had decreased from 51% in 2005 to 44% in 2010 but has stayed almost the same at 43% in 2012 [4].

In addition, according to the results of the research conducted by Ministry of Health (MOH), Ministry of Agriculture and Livestock Resources (MINAGRI) and Ministry of Local Government (MINALOC) in 2013, it was found that 38% of children were stunted (below -2 SD where SD means Standard Deviation), and 14% were severely stunted (below -3 SD). Stunting increased with age, peaking at 49% among children age 18-23 months. A higher proportion of male (43%) than female (33%) children were stunted. Stunting affected children in the rural areas (41%) more than those in the urban areas (24%). Stunting was inversely correlated with the mother's education level and household wealth quintile. For example, 49% of children in the lowest wealth quintile were stunted, as compared with 21% of children in the highest quintile [1].
In the same research, 2% of children were found wasted and the percentage was higher in rural areas than urban areas. The same results showed that 9% of all children were underweight, and 2% were severely underweight even though there was no variation of underweight by sex of the child.

The proportion of children who were underweight was greater in rural areas (10%) than urban areas (6%). Moreover, this was inversely correlated with the mother's education level.

Anemia is also a major issue from malnutrition. It was found that 37% of all consulted children suffered from some degree of anemia where 21% were classified as mildly anemic, 15% were moderately anemic, and less than 1% was severely anemic. This decreased with age but no significant difference in girls and boys.
The improvement in nutrition is justified by poverty reduction where poverty reduced from 44.9% in 2011 to 39.1% in 2014 and extreme poverty from 24.1% to 16.3%. Inequality reduced as well since the Gini coefficient has dropped from 0.49 in 2011 to 0.45 in 2014 [5]. Gini coefficient is a measure of inequality which is expressed in percentages or most of times in decimals taking values between 0 and 1. Note that 0 stands for perfect equality and 1 for perfect inequality [5].
According to the results from different Rwanda Demographic and Household Surveys (RDHSs) in the years 2005, 2010 and 2014/15 [5]; the corresponding respective statistics found are given in following table:

	RDHS(2005)	RDHS(2010)	RDHS(2014/15)
Stunting	51%	44%	38%
Wasting	5%	3%	2%
Underweight	18%	11%	9%

Another programme called in-home fortification and nutrition education to combat anemia and micronutrient deficiencies among children 6-23 months in Rwanda was settled and it had impact on households living conditions. The critical developmental of the brain, motor skills and social-emotional skills period is often referred to as the 1000 days, which includes the time from conception together with the two first years of life. Mothers levels of education and knowledge on health are related to stunting [8].

The use of micronutrient powders (MNP) mainly iron was found useful in the Worldwide for the in-home food fortification mainly for children. Children who suffer from deficiencies of micronutrients early in life, particularly iron and iodine, are at higher risk of suffering from irreversible impairment of physical and cognitive development.
However, only 20 percent of Rwandan children consume food rich in iron [3]. According to the recent results of a cost of hunger study in Rwanda, 21.9 percent of child mortality is associated with undernutrition [3].
To strengthen the efforts made in the fight against malnutrition in Rwanda, Small Area Estimation technique may help us to investigate the most needy Districts in terms of nutrition.

1.2. Problem Statement

It is a general remark that Rwanda still needs much efforts to eradicate malnutrition. Referring to the percentage of malnutrition obtained in previous years, can it be assumed to be the same in all districts in the country? In other words; are all districts likely equally suffering from malnutrition? The answer may be NO. Therefore; it is better to find the appropriate statistical method or technique to be used in order to find the proportions of malnutrition in each of the thirty districts of Rwanda. Small Area Estimation technique is recommended to find the characteristics of subdomains, which in our case, will be Districts. Before all; we will have a review on this technique in order to understand it and how it works. The goal is to know the neediest districts so that policy-makers may not be misled mainly when allocating funds.

1.3. Indicators of malnutrition

a. Stunting (Height for Age) reflects chronic undernutrition during the most critical periods of growth and development in early time. A child aged 0 to 59 months whose height for age is below minus two standard deviations from the median of the World Health Organization (WHO) Child Growth Standards is said to be moderately and severely stunted. If the height is minus three standard deviations, the child is said to be severely stunted. Stunting refers to the skeletal growth. Stunting is also called shortness which means low height relative to age.

b. Wasting (Weight for Height) reflects acute undernutrition. Wasting refers to recent weight loss or gain. The measures are the same as for Stunting in the respective categories. Wasting is also called thinness which means low body weight relative to height.

c. Underweight (Weight for Age)is a composite form of undernutrition that includes elements of stunting and wasting. The measures are the same as for Stunting in the respective categories. Underweight is as a combination of both stunting and wasting and then, it does not distinguish between acute malnutrition (wasting) and chronic malnutrition (stunting). Children can be underweight for their age because they are stunted, wasted, or both. Weight-for-age is an overall indicator of a population's nutritional health.

d. Severe acute malnutrition is defined as the percentage of children aged 6 to 59 months whose weight for height is below minus three standard deviations from the median of the WHO Child Growth Standards, or by a mid-upper-arm circumference less than 115 mm, with or without nutritional oedema.

e. Overweight is said when a child aged 0 to 59 months whose weight for height is above two standard deviations (overweight and obese) or above three standard deviations (obese) from the median of the WHO Child Growth Standards.

f. Low birthweight is defined as a weight of less than 2,500 grams at birth. This is associated with poor nutrition in mothers [6].

Stunting manifests itself after a long period of time nearly after 2 years of age and is also called shortness. In reverse its recovery is very hard and can take so long time. Wasting manifests itself commonly in infants and younger children often during the weaning period and is also called thinness. As it is from malnutrition of a short recent period of time, it can be recovered.
The chronic hunger is the status of people receiving their regular food intake which provides them less energy than required. This leads to undernutrition.
In this study, we will use Elbers, Lanjouw and Lanjouw method also called the World bank method to find district estimates of proportions of malnutrition in all its aspects. In other words, we will estimate the proportions of the three indicators of malnutrition in each district.

Districts proportions are needed in order to help in decision making by policy-makers because we cannot expect all districts to be likely equally suffering from malnutrition. Importantly in funds allocation, it is better to focus on districts found to be the most suffering. In absence of such information, the funds can be misused due to the uniformity assumed when distributing some specific resources. It is known that life conditions change region by region or domain by domain. This makes small area (district in our case) estimates needed to be efficient in the decision and policy-making.

2. Small Area Estimation (SAE) methodology

2.1. Motivation and background

In our daily life, we need planning for it to be better. Planning consists of making a set of decisions to be implemented in the future. Those decisions are drawn from observations of the present and the past. Here, we introduce the concept of research. To have information about some variables of interest, we conduct an appropriate research.
It is preferable to conduct a census over the whole population but it costs much so that it cannot be afforded whenever needed. Whatever we do is valued in terms of time. Time constraint is also a major problem which makes us conduct surveys rather than censuses to get information on the population. The enumeration costs much due to all expenses allocated to it. Here, we highlight the accessibility of data from respondents involved in the research. As the number of respondents increases, the enumeration becomes harder. The displacement to arrive to everyone in the target population is expensive of course. In addition to the enumeration, we have all preparation of the research before all, analysis of data and their interpretation from which we deduce the recommendations to policy-makers.
It is common that many times, we meet the budget constraint. However, time constraint will also be substantial in the research as we always plan considering the scope of time referring to when the results are needed.
Surveys are solutions to this issue. We do not always need to find data from the whole population but we can consider a part of the population to get information and then we make a generalization over the whole population. Note that the sample must be representative of the whole population.
From the population units (e.g.,people, objects); we select some of them so that the results from studying the selected units can be fairly generalized to the whole population. This process of selection of a sub-population is called **Sampling**. The selected sub-population is called a **Sample**.
The intended sample must be such that the conclusions deduced from it allow us to make inference to the whole population. In this case, the sample is said to be representative. This is because it represents the population in terms of information, needs and all aspects.
In general, we get the population information referring to the sample from it. Sampling is a double-interest work where we aim at the reduction of both time and money but always arriving at adequate estimates of parameters of interest. Note that a sample can be chosen from the whole population or a sub-population which can be a specific region.

But surely, we will have some errors because all elements of the population in the research are not taken into account.
In other words, a great part of errors is due to the fact that we take a part of

the population but not the whole population. The success is the minimization of errors such that the obtained estimates are of adequate precision.
It would be more efficient if we had used the information received individual-wise (each individual could be considered alone which is often difficult. For the same variable of interest, all the population will not be homogeneous. The generalization of conclusions over the population in a specific region is different from the generalization over the whole country. When we consider the whole population, we can have a diversity of sub-populations where these last can be nearly similar or totally different depending on the characteristics taken into account.

The presence of information on specific domains of the population became important in decision-making instead of the general information where policies could be inappropriate for some domains. For the accuracy, we need some estimates for a specific region or sub-population rather than the whole country. It is not logical to take the same decision over the whole country when dealing with some characteristics of the population.

In this work, we consider SMALL AREA ESTIMATION method where we need to find the estimates of some characteristics of interest for sub-populations. The population is divided into sub-populations also called domains. These last can be geographical, socio-economical or even socio-demographic. In the same way a domain can be subdivided into sub-domains. The estimation of parameters will depend upon the level of interest if it is either the sub-domain or a domain. A domain is called a small area if the sample size of the sample drawn from it is so small that it produces the estimates of low precision. Small sample sizes are the main causes of biasedness of estimates in terms of the variability. The main purpose can be of finding the estimates at the lowest level/subdivision of the domain or at the domain level.
As explained above, the sample of a given area can be large enough to produce good estimates. In other words, among possible causes of this problem, there is the small sample size. If the estimates obtained from a sample of a domain allow us to make inference to the population, the area is not said to be small. On the other hand, a domain is said to be small if the estimates produced by its sample are not of adequate precision that they can be used to make inference for the population. In this case, we can either use some other variables also called auxiliary variables (covariates) or auxiliary information if available or direct design-unbiased estimation by randomization distribution. (The issue could also be caused by bad sampling frame where it becomes not easy to get a representative sample. Model diagnostics can be a solution to model-based issues arising from prediction model misspecification.

The auxiliary variables are used to strengthen the estimates which were found to be of inadequate precision. The auxiliary variables must be of high prediction power (to be linked to the variable of interest).
As said above, the problem can also be from the fact that there has not been an appropriate model to predict the estimate at a desired precision. Therefore, the model assumptions and model diagnostics are of great importance because the model specification is also a major task in prediction. In addition, the estimates cannot be efficient when the variable is highly correlated with others that would be used to explain more the variation/ changes in the variables of interest. Hence; small areas are those with small sample sizes which need borrowing strength from other data related to the area in one way or another. All these techniques combined to produce estimates of the population with adequate precision is called :**Small Area Estimation**(SAE)[9].

When selecting the sample of the overall population, we get elements in different domains (areas). Since the sampling is random, some domains will be poorly represented (to mean that their sample sizes are very small) in the overall sample that their specific direct estimates will not be adequately precise. Such domains are also called: sub-domains, minor domain, local area or small area, small subgroups or sub-provinces.

2.2. Interest of the topic

The greatest error which is made by people is the generalization in their conclusions after a given number of observations which is not enough. In most researches, the conclusions are drawn from the sample and then inferred to the whole population. These inferences of population statistics from the sample statistics can make decision-makers fall in mistakes if there has not been paid much attention in parameters estimation. Here, small area estimation method is used to find the estimates for small areas where the sample is not representative or does not allow us to draw reliable conclusions for the domain.

In other words; it is not very reasonable to take decision referring to the sample statistics without looking at their precision. Most of cases will be that the sample was selected at an aggregate level where a specific domain cannot probably be represented as desired. For more precision, it is better for the sample selected purposely to the domain where even its size is considered at the beginning. For example, if you consider the CPI (Per Capita Income); you can attribute the same CPI to two persons abiding different and spaced areas but it is known that the average CPI for cities (or urban areas) and villages (or rural areas) will be different.

From the randomness of the sample, we can have many elements in one sub-domain rather than another. The sampling distribution will be of great role in the adequacy of estimates. But; at least if the two persons are in the same sub-domain; their CPI can be (neighbors) oscillating around a given amount with small difference.

Probably; we will have the diversity in the population of a wide domain compared to a small domain. This can be explained by the fact that people of the same social classes (socio-demographic or socio-economic) tend to abide the same area. It is therefore wise to estimate the statistics of an individual referring to the domain he/she is included in but having dealt with the domain representation in the sample of the whole population.

Small Area Estimation technique is useful for Governments and non-governmental organizations (NGOs) mainly in funds allocation; regional policy-and decision-making and business planning. It is very much used for the social and economic policy-making. The efficiency of this is that the efforts to be made in the development of a specific region/area or domain are different to those to be made in another. For example; when allocating heath funds, disease prevention funds or poverty reduction funds; we will take into account the neediest sub-populations and we assign the appropriate proportion to each specifically because all domains are not in need equally. Then, we avoid the error of generalization which is not relevant even if the common remark is that a sub-domain contains many aspects of the domain containing it. This is why we prefer using the data at the aggregate level to strengthen the results which would be obtained from the sample of the lower level. Of course, there is a kind of homogeneity of units in the same domains for some characteristics.

Therefore; the data from the population survey or census will play a great role in the estimation of the small area parameters/estimates. This is, with population data and sample survey, we will obtain very good estimates for the small area. Here, we do not believe the representativity of the sample in the population. This is, when sampling in the whole population, all the areas are not well/ sufficiently represented.
Then; we can be mislead by the estimates of the population when considering the sub-populations. But; because some of the elements in the sub-population were selected in the sample population; this is why we do not reject the population estimates because they will help us to be more accurate in the estimation. Moreover, when we are dealing with the small area estimation, we use auxiliary information. This is the information we draw from other (preferably neighboring) areas, the same area in the previous time or the whole population census.

The auxiliary data are those which can also help us to get some information on the target variable. Of course, they can be linked to the target variable in one way or another. The population parameters can give us information on the specific domain but not fully. This is why when making a business decision, we can advise someone to use the small area estimation technique in the market study. This will help you to allocate your business bureau, stations (activities) appropriately. It is the same as for the educational or health programs allocation as said above. Note that the small area estimates are useful to both private and public sector.
In applications, it can happen that the sample size is zero in a given domain, the case when we are advised to refer to or use the auxiliary data. We note that the sample size for the area is the major cause of reliability of its direct estimates. The greater the sample size, the more reliable the domain estimates. This is from the fact that the sample size influences the variance of the estimates and to avoid large variance, we consider the large sample size.

2.3. Users of small area estimation

Small area estimation is used in socio-demographic, socio-economic studies. It is used in poverty mapping, disease mapping where we investigate the most suffering or the most exposed people, regions either from a disease or poverty. For example, it was used in poverty mapping in Philippines, Indonesia,etc. It is used to investigate the Unemployment rates according to regions or social classes. It is also used in demography where we need to study the population movements, changes or simply population dynamics.

3. Small Area Estimation Approaches

The estimate obtained immediately using the sample data is called: DIRECT ESTIMATE. In the opposite case, we obtain INDIRECT ESTIMATE. This is an estimate obtained from the data of the small area of interest as well as from other areas and other variables, or even from other data sources (other surveys, registers or censuses). These data can be the data from other surveys done before or surveys done in neighboring areas and in most cases, they are from local administrative records.
The name indirect is due to the fact that these estimates use data from other areas (domain indirect estimates) or time (time indirect estimates) or both (domain and time indirect estimates).
Therefore, the small area is the area or sub-domain which will need the intervention of covariates to produce reliable estimates. Reliable estimates are the estimates with acceptable precision. In this case, indirect estimate will be used instead of direct estimate. Let us consider the variable of interest Y with values y_{ij} and the auxiliary variables X with values x_{ij} for the j^{th} unit in the i^{th} small area.
Researchers can be interested in finding Totals, Means or Proportions. Throughout this thesis, we will consider the case when we need the mean of the target variable. In many cases, a set of the means of the auxiliary variables is used at the area level which is written as $\overline{X}_i$

3.1. Efficiency of an estimator

For the test of the precision and accuracy of an estimator, we refer to its Mean Squared Error (MSE) which increases with the variability and biasedness of the estimator. MSE of an estimator gives us information on the precision and accuracy of an estimator which are its two main qualities.
The greater the MSE, the poorer the estimator. An estimator is intended to be both precise and accurate but this is not always the case. There can be cases where it is either precise but not accurate or accurate but not precise. Let us consider that we are still interested in estimating θ (which is the parameter of interest).

$$MSE(\widehat{\theta}) = Var(\widehat{\theta}) + [Bias(\widehat{\theta})]^2 \tag{1}$$

Since $MSE = E[(\widehat{\theta} - \theta)^2] = E[(\widehat{\theta} - E(\widehat{\theta}) + E(\widehat{\theta}) - \theta)^2] = E[(\widehat{\theta} - E(\widehat{\theta}))^2 + E(E(\widehat{\theta}) - \theta)^2] + 2E[[\widehat{\theta} - E(\widehat{\theta})][E(\widehat{\theta}) - \theta]]$ where the last expression in product becomes zero and with $Var(\widehat{\theta}) = E[(\widehat{\theta} - E(\widehat{\theta}))^2]$ and $Bias^2(\widehat{\theta}) = E[(E(\widehat{\theta}) - \theta)^2]$.
We remark that if an estimator $\widehat{\theta}$ of θ is asymptotically unbiased (this is, $Bias^2(\widehat{\theta}) \to 0$) and remain with great MSEs, it means that it has large variance.

In case of choice, an unbiased estimator with large variance is better than a biased estimator with small variance. This last is precise but not accurate (far from the real value) while at least the first can help us to estimate the confidence interval of our estimator.
In the case of indirect estimators; the time indirect information is better than the domain indirect information because time series data will tend to give the same indication or conclusion over the same variable of interest unless the structure of the population has changed over time. However, the weighted linear combination of the two is the best rather than either indirect information. The MSE of the weighted linear combination of two estimators is expected to be smaller than the MSE of either component estimator.
This is due to the fact that the estimator with high variability will have a small contribution while the estimator with low variability will contribute much to that linear combination. MSE is the sum of the variance and the bias of an estimator.
The greater is the MSE the weaker the estimator. The estimator with high MSE is thought to be biased and its bias can be explained by that great amount of MSE. In our estimation, we aim at finding good estimators. A good estimator is the one which is accurate and precise. Based on confidence intervals, the variability of an estimator can cause decision makers and policy makers commit mistakes. If the confidence interval is wide, it will be difficult to take appropriate decisions.

The width of the confidence interval will give us information on the precision of our estimator. This lack of precision observed from the width of the confidence interval is linked to the variability of an estimator due to the fact that in the construction of the confidence interval of an estimator, its variance plays an important role. If both the estimator and its (design-)variance are p-consistent, the confidence interval will contain the true value of the population. All these problems are due to the fact that the data we use are not for the whole population where every unit in the population is consulted.
This is why we need to choose the sample size under condition that we do not exceed the desired tolerances on the Coefficients of Variations (CVs) of the direct estimators. The coefficient of variation (CV) is also referred to as an indicator of variability of an estimator. When looking at the variability of an estimator, we consider the coefficient of variation (CV) which is expressed in terms of percentages. In addition, the CV is more easily understandable than MSE.
In the case of direct estimation, we have:

$$CV(\overline{\theta}) = \frac{s(\overline{\theta})}{\overline{\theta}} \tag{2}$$

where $s(\overline{\theta})$ is the standard error.
There are two main approaches which are used in Small Area Estimation where we have Design-based approach and Model-based approach.

3.2. Design-based approach

3.2.1. Direct estimator

A direct estimator is an estimator obtained using only from the sample data of the area and the variable of interest. Denote by $N_i = \sum_{i=1}^{m} n_i$ the overall sample size in all m areas where n_i is the sample size corresponding to the i^{th} area. Direct estimators are usually unbiased, though they may have large variances. Suppose that our interest is the area mean ($\overline{Y}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} y_{ij}$). The area mean direct estimator is given by:

$$\widehat{\overline{Y}}_{i,D} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} \tag{3}$$

where n_i is the sample size in the i^{th} small area and y_{ij}'s are the observations of the variable of interest Y for the selected sample [9]. The index D on the variance Y_i shows that we have the direct estimator. In other words, the direct estimator of the area mean is the sample mean. This is why the sample must be representative in order to have no lack of information.

The estimator depend always on the sampling design which was used. If we used the simple random sampling without replacement, the direct estimator will have the conditional design variance given by:

$$V_D(\widehat{\overline{Y}}_{i,D}) = V_D(\overline{y}_i|n_i) = (1 - \frac{n_i}{N_i})\frac{S_i^2}{n_i} \tag{4}$$

where the quantity $S_i^2 = \frac{1}{N_i-1}\sum_{j=1}^{N_i}(y_{ij} - \overline{Y}_i)^2$ has as unbiased estimator $s_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_i)^2$ where $\overline{Y}_i$ and $\overline{y}_i$ are the total area mean and sample area mean respectively [16].
Therefore, the variance of the estimator becomes:

$$V_D(\widehat{\overline{Y}}_{i,D}) = V_D(\overline{y}_i|n_i) = \frac{N_i - n_i}{n_i N_i(n_i - 1)}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_i)^2 \tag{5}$$

Substituting S_i^2 by its unbiased estimator s_i^2 in (4); one gets (5).
Notice that $V_D(\overline{Y}_{i,D}|n_i)$ is $O(1/n_i)$ and hence becomes small for n_i which is large. The index D on the variance $V_D(.)$ shows that we have the design variance [13],[16].
If we have a set of covariates for the sample, we can obtain the design-based estimator which will be more efficient compared to the previous direct

estimator. This is called the regression estimator. It will be of the form:

$$\widehat{\overline{Y}}_i^{regr} = \overline{y}_i + (\overline{X}_i - \overline{x}_i)' \beta_i \tag{6}$$

where $\overline{X}_i$ and $\overline{x}_i$ are the vector means of covariates for the whole small area and the corresponding sample respectively. β_i is a vector of regression coefficients (specific) for the area. Its design variance will be of the form: $V_D(\widehat{\overline{Y}}_i^{regr}) = (1 - \rho_i^2) V_D(\widehat{\overline{Y}}_{i,D})$ [13].
According to a sampling design used, we get a number k of possible samples s_i's each with a specific probability $p(s_i)$ to be chosen. Let $\overline{\theta}$ be the estimate of the the parameter of interest θ. Then $\overline{\theta}$ is design-unbiased (also called p-unbiased)if its design expectation noted as $E_p(\overline{\theta})$ is equal to θ. That is,

$$E_p(\overline{\theta}) = \sum_{i=1}^{k} p(s_i) \overline{\theta}_i = \theta \tag{7}$$

where the θ_i is the estimate in the i^{th} sample (or in the sample s_i) and $p(s_i)$ is its probability to be chosen. The corresponding design variance is:

$$V_p(\overline{\theta}) = E_p(\overline{\theta} - E_p(\overline{\theta}))^2 \tag{8}$$

with estimator $\upsilon(\theta)$ which becomes p-unbiased when $E_p(\upsilon(\theta)) = V_p(\overline{\theta})$. If the p-bias and design-variance of an estimator $\overline{\theta}$ tend to zero as the sample size increases, then $\overline{\theta}$ is said to be design-consistent or p-consistent. When simply the p-bias tends to zero, the estimator is said to be p-unbiased [9].

3.2.2. Direct Synthetic estimators

We consider the case where we have m areas with sample size n_i for the i^{th} area. When we are interested in estimation of the population mean, we use the sample of size n where $n = \sum_{i=1}^{m} n_i$. This is, the overall sample size is the sum of the sample sizes of the areas. For the i^{th} area, we get the area sample mean as $\overline{Y}_{i,s} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ for $i = 1, 2, ..., m$. The direct synthetic estimator (which will be denoted by $\overline{Y}_S$) is given by the relation below:

$$\overline{Y}_S = \frac{1}{\sum_{i=1}^{m} n_i} \sum_{i=1}^{m} n_i \overline{Y}_{i,s} \tag{9}$$

The composite estimator is:

$$\overline{Y}_{i,C} = w_i \overline{Y}_{i,s} + (1 - w_i) \overline{Y}_S \tag{10}$$

where $w_i = \frac{n_i}{N_i}$ or $w_i = \frac{\sum_{i=1}^{m} n_i}{\sum_{i=1}^{m} N_i}$ [9].

3.2.3. Ratio synthetic estimator

In presence of auxiliary information, we get the synthetic estimator. We let $\overline{X}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} x_{ij}$ and $\overline{x}_{i,s} = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ where s indicates that the mean if from the sample. The ratio synthetic estimator will be given by:

$$\widehat{\overline{Y}}_{i,RS} = \frac{\overline{y}_s}{\overline{x}_s}\overline{X}_i \tag{11}$$

where $\overline{y}_s$ and $\overline{x}_s$ are from the overall sample of size $n = \sum_{i=1}^{m} n_i$. This is,$\overline{y}_s = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}$ and $\overline{y}_s = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij}$ [9],[21]. However, we could use the sample from the area and the ratio synthetic estimator above will be of the form: $\overline{Y}_{i,RS} = \frac{\overline{y}_{i,S}}{\overline{x}_{i,S}}\overline{X}_i$. The composite estimator will be obtained by using the mean of the auxiliary information of the non-sampled items in the area and will take the following form:

$$\overline{Y}_{i,C} = w_i\overline{Y}_{i,s} + (1-w_i)\frac{\overline{y}_s}{\overline{x}_s}\overline{X}_i^{'} \tag{12}$$

where w_i is as defined previously and $\overline{X}_i^{'} = \frac{1}{N_i - n_i}\sum_{j=n_i+1}^{N_i} x_{ij}$.
The used ratio $\frac{\overline{y}_s}{\overline{x}_s}$ is preferable and this is explained by the fact that it is the one which minimizes the standard residuals. We first assume the following distribution:
$y_{ij}|\mu \sim indN(\mu x_{ij}, \sigma^2 x_{ij})$ and we aim at μ minimizing the quantity:
$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \mu x_{ij})^2/\sigma^2 x_{ij}$ which is equivalent to: $\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}^2/x_{ij} - 2\mu y_{ij} + \mu^2 x_{ij})$.
This is to mean that we are finding the critical μ for the minimum of the function $f(\mu) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}^2/x_{ij} - 2\mu y_{ij} + \mu^2 x_{ij})$ for all x_{ij} and y_{ij}. In other words, since σ^2 is a constant and using the derivative with respect to μ to find the optimum, we have the following:

$$min(f(\mu)) = min(\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}^2/x_{ij} - 2\mu y_{ij} + \mu^2 x_{ij}))$$

$$\mu\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij} - \sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij} = 0$$

$$\widehat{\mu} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij}} = \frac{\overline{y}_s}{\overline{x}_s}$$

which is the recommended ratio.
Note that the auxiliary variables X_{ij} are collected in the sample for the sampled units while for the non-sampled units, we use those from the recent population census.

For each area $i = 1, ..., m$, we have the conditional expected mean of the characteristics of interest as below:

$$E[\overline{y}_i|y_{i1}, y_{i2}, ..., y_{in_i}] = w_i\overline{Y}_{i,s} + (1 - w_i)\mu\overline{X}_i^{'} \tag{13}$$

where μ is substituted by its estimate $\widehat{\mu}$.
Therefore, we get the following distribution $y_{ij}|\mu \sim indN(\mu x_{ij}, \sigma^2 x_{ij})$ where $x_{ij}(> 0)$ and σ^2 are known. In addition, μ is also uniformly distributed on $\mathbb{R}$.
Moreover, μ is such that $\mu|y_s \sim N(\frac{\overline{y}_S}{\overline{x}_S}, \frac{\sigma^2}{n\overline{x}_S})$ where $n = \sum_{i=1}^{m} n_i$.
For non-sampled items, the joint posterior distribution of the characteristic of interest conditional to synthetic estimator $\overline{y}_s$ is a multivariate normal distribution obtained as: $y_{ij}|y_s \sim MVN(\frac{\overline{y}_s}{\overline{x}_s}x_{ij}, \frac{\sigma^2}{n\overline{x}_s}x_{ij}^2 + \sigma^2 x_{ij})$.
Their covariance will be $cov(y_{ij}, y_{i'j'}|y_s) = \frac{\sigma^2}{n\overline{x}_s}x_{ij}x_{i'j'}^{'}$ where $x_{i'j'}$ and $y_{i'j'}$ are the characteristics of non-sampled items for another area $i^{'}$ different from the area i.

3.2.4. Indirect estimators

As explained at the beginning of the Section, indirect estimates are used when the direct estimates are not precise enough to allow us to make inference over the whole population. An estimate can be direct or indirect depending on what is taken into account where we can consider time, domain or both as explained previously. Indirect estimators are based on implicit or explicit models that incorporate the available information. For example, information obtained in a survey can be combined with the one collected in a census or an administrative register. Indirect estimators are usually biased, although their variances are smaller than for the direct (unbiased) estimators, and the trade-off of bias and variance is usually in their favor. Auxiliary variables are also called covariates. The efficiency of indirect estimators depends upon two elements:

a. Availability of auxiliary data which are appropriate,

b. Good specification of the linking model.

To say that the auxiliary data are appropriate means that they are in one way or another related to the variable of interest. If there is no correlation between them, the corresponding estimate will remain inefficient. If there exists correlation between the two, the estimate will be more precise. In this case, the auxiliary variable are said to be of good prediction power. It can happen that the auxiliary data can be used alone to estimate the target parameter without sample data but the efficient case will be when we combine both data.

The choice of the linking model is also important task in prediction since there are errors related to model misspecification. Such errors are called model errors. Appropriate linking model will help us in the errors reduction. It is possible that lack of precision of estimates is due to bad choice of the model.
The two tasks (choice of appropriate linking model and auxiliary variables of good prediction power) will be of great importance (for the success) in the estimation of indirect estimates.

3.2.5. Synthetic estimators

For synthetic estimates, we assume that a set of direct estimates of larger domains of the population is available. This works under the assumption that the area has the same properties as the larger domain containing it and whose direct estimates are available/known. This estimator is called synthetic because it is obtained with the use of an estimate of the domain covering many small areas assumed to be homogeneous. Of course, the synthetic estimators borrow information from other similar areas. Let us denote the variable of interest by Y where Y_{ij} denotes its value for the j^{th} individual(item) in the i^{th} area. The unit level model will be:

$$Y_{ij} = X_{ij}^{'}\beta + u_i + e_{ij} \tag{14}$$

where u_i and e_{ij} denote the area-specific effects and the unit-specific effects respectively whose means are expected to be zero [9]. X_{ij} is a set of auxiliary data for the auxiliary variables available at unit level. β is the vector of parameters of estimation relating the target variable Y to the set of auxiliary variables X_{ij}'s correspondingly. In the equation above, $X_{ij}^{'}$ is the transpose of X_{ij}. Y_{ij} and X_{ij} indicate the values of the target variable Y and auxiliary variables X's, respectively, where i $(i = 1, 2, ..., m)$ stands for the area and j $(j = 1, 2, ..., n_i)$ stands for the j^{th} unit in the given area. The equation above is valid in the case of the availability of auxiliary variables at unit level.
In this case, the small area mean $\overline{Y}_i$ of the target variable Y is obtained by:

$$\overline{Y}_i = \overline{X}_i^{'}\beta + u_i \tag{15}$$

where $\overline{X}_i$ is a set of mean vector of the available auxiliary data (variables). Using least square errors method,we obtain the estimated value of β and the expected synthetic area mean estimate becomes as follows:

$$\widehat{\overline{Y}}_{i,S} = \overline{X}_i^{'}\widehat{\beta} = \frac{1}{N_i}\sum_{j=1}^{N_i} x_{ij}^{'}\widehat{\beta} \tag{16}$$

The ordinary least squares estimator $\widehat{\beta}$ of β is obtained from the relation:

$$\widehat{\beta} = [\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij}x_{ij}^{'}]^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij}y_{ij}^{'} \tag{17}$$

where m is the total number of areas and n_i is the sample size for the i^{th} area [16].
The synthetic estimator is more advantageous than the direct estimator because we use β which is estimated by using data of a large sample while for direct estimator, β was estimated by using area sample data. That is, for synthetic estimator, $V_D(\overline{Y}_{i,S}) = O(1/n)$ with $n = \sum_{i=1}^{m} n_i > n_i \ \forall \ n_i$ and then the design variance reduces. This is to mean that the overall sample size n, is obtained by summing up the partial sample sizes in all the m areas. In addition, the motivation of using the synthetic estimators is from the fact that the regression parameters vector β is common to all areas. This is to mean that the bias will increase when we use different β_i's for different areas instead of one common β [13].

3.2.6. Sampling with unequal probabilities

It is not common that we will always deal with the random sampling where all elements are equally likely to be selected in a sample. We now see the case when each element in the population has its probability to be selected. We define w_{ij} to be the weight of the j^{th} element in the i^{th} area where w_{ij} is the inverse of the probability of the element to belong to the selected sample s. That is, $w_{ij} = 1/p[(i,j) \in s]$
Coming back to the case when we are interested in the area mean, we will obtain:

$$\widehat{\overline{Y}}_{i,D} = \frac{\sum_{s_i} w_{ij}y_{ij}}{\sum_{s_i} w_{ij}} \tag{18}$$

where s_i indicates the sample in the i^{th} area and $\sum_{s_i}$ indicates the sum over the sample. This is, the sum is found over the selected items in the area sample s_i since $\widehat{\overline{Y}}_{i,D}$ represents the area mean direct estimate. Notice that the w_{ij} is the weight of the j^{th} element in the sample of the i^{th} area and is interpreted as the number of elements in the same area represented by the j^{th} selected element [10].

For the regression synthetic estimator in this case, the resulting regression coefficients vector will be called **probability weighted** estimator β_{PW} instead of ordinary least squares estimator β_{OLS} used previously and it is

given by:

$$\widehat{\beta}_{PW} = [\sum_{i=1}^{m}\sum_{j=1}^{n_i} w_{ij}x_{ij}x_{ij}^{'}]^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i} w_{ij}x_{ij}y_{ij}^{'} \tag{19}$$

Since the synthetic estimator are expected to have large bias, we try to substract this last from the estimator and we get the following survey regression estimator:

$$\widehat{\overline{Y}}_{i,S-R} = \overline{X}_i^{'}\widehat{\beta}_{PW} + \frac{1}{N_i}\sum_{j=1}^{n_i}(y_{ij} - x_{ij}^{'}\widehat{\beta}_{PW}) = \widehat{\overline{Y}}_{i,H-T} + (\overline{X}_i - \widehat{\overline{X}}_{i,H-T})\widehat{\beta}_{PW} \tag{20}$$

where the index $H-T$ indicates the Horvitz-Thompson estimator [12].

3.2.7. Composite estimators

A composite estimator is a weighted average of direct estimate and regression synthetic estimate. Since it is a linear combination of the two estimates; it permits the trade-off between their advantages and disadvantages. Having chosen appropriate weights, the synthetic estimate must have smaller square mean errors (SMEs) compared to those of the either components. Therefore, the selection of the weights will be an important task but not easy!
We consider the sampling model which gives the relation between the area-specific parameter and the expected population parameter with sampling variance assumed to be known. The linking model is the model expressing the expected population parameter in terms of model parameter β with the model variance. The combination of the two models will be a special case of the linear mixed model.

For the composite estimator, more weight is given to the component estimator with small variance. Therefore, if the sampling variance is large, more weight is given to the synthetic regression estimator. The model parameters can be estimated by use of moment method (MM) by Fay-Herriot (1979); Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) methods. The obtained estimator is called Empirical Bayes or Empirical Best estimator.
We can test the significance of the model variance and see if it can be approximated to zero. If this hypothesis is not rejected, we can use only the synthetic estimator with no random effects. This probably will happen when the selected auxiliary variables (covariates) are much correlated with the variable of interest. This is why the choice of covariates is also an important work which will contribute to the achievement of our goal. The goal is to obtain an unbiased estimator.
In the same way, we can prefer using only the direct estimator when the

model variance is found large enough. However, this is why we prefer the weighting method (weighted linear combination) where each component is involved in the estimation of the composite estimator but at different proportion in accordance with the precision of each. Hence, the composite estimator $\overline{Y}_{i,C}$ for the i^{th} area will be given by:

$$\overline{Y}_{i,C} = \gamma \overline{Y}_{i,D} + (1-\gamma)\overline{Y}_{i,S} \tag{21}$$

where $\overline{Y}_{i,D}$ and $\overline{Y}_{i,S}$ represent the area mean direct estimator and synthetic estimator, respectively. The weight is determined by the ration of MSEs such that the estimator whith high MSE contribute less(since it is less precise): We have γ of the form:

$$\gamma = \frac{MSE(\overline{Y}_{i,S})}{MSE(\overline{Y}_{i,S}) + MSE(\overline{Y}_{i,D})} \tag{22}$$

where the $MSE(\overline{Y}_{i,D})$ and $MSE(\overline{Y}_{i,S})$ are the Mean Squared Errors of the direct estimator and synthetic estimators respectively [9],[12]. Looking at the expression of γ, we remark that if $MSE(\overline{Y}_{i,S}) > MSE(\overline{Y}_{i,D})$ then much weight is attributed to the direct estimator $Y_{i,D}$ and vice versa.
Since MSE is the sum of variance and bias, the remarks are the same when using variances of both estimators.

3.2.8. Generalised Regression estimator

Let us be interested in finding the area totals which will be denoted by Y in an area under consideration. In presence of auxiliary information which is a vector of totals, $X = (X_1, X_2, ..., X_p)^{'}$ for p variables. For the sample which was chosen, we get the observations in the form: (y_j, x_j) where x_j is the values of the auxiliary variables for the j^{th} unit in the sample from the area.
The generalized regression estimator $\widehat{Y}_{GREG}$ is expressed by:

$$\widehat{Y}_{GREG} = \widehat{Y} + (X - \widehat{X})^{'}\widehat{\beta} \tag{23}$$

where $\widehat{X} = \sum_{j\in s} w_j x_j$ and $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2, ..., \widehat{\beta}_p)$ is obtained by weighted least squares from the sample observations.
That is, $\widehat{\beta} = (\sum_{j\in s} w_j x_j x_j^{'}/c_j)^{-1} \sum_{j\in s} w_j x_j y_j/c_j$ with c_j's the specified positive constants [9].

3.3. Model-based Approach

This approach assumes a model for sample data and use the (approximately) optimal predictor of the target variable. What is obtained is a predictor

(whose values are random) because we assumed the model. Notice that the accuracy of predictors is still a challenge in the presence of small sample size or covariates with low predictive power. It is also probable to have the problem of model misspecification. Common used models are classified into two types of models depending on the level of available auxiliary data where we have:

a) Unit level models

b) Area level models

3.3.1. Unit level model

As said above, this type of models is used if the unit level auxiliary information is available. We consider the case when we have n individual-level observations of the variable of interest Y. That is, $(y_{i1}, y_{i2}, ..., y_{in})$ and the corresponding (chosen) auxiliary variables x_{ij} whose area means $\overline{X}_i$ are known.
The unit level model is also known as Nested Error Regression Model. It will have unit level random effects and area level random effects. We consider the case when we have a set of auxiliary variables $\mathbf{x}_{ij}$ for the j^{th} unit in the i^{th} area. Since we are aiming at finding the area mean $\overline{Y}_i$, we use the means $\overline{\mathbf{x}}_i$ or totals $\mathbf{x}_i$ of those auxiliary variables accordingly. But, in this case, we are estimating the unit level value Y_{ij}. Then, the Nested Error Regression Model will be written as follows:

$$Y_{ij} = x_{ij}^{'}\beta + u_i + e_{ij} \tag{24}$$

where random area effects u_i's are independent and identically distributed with mean 0 and variance σ_u^2 (this is, $u_i \sim iidN(0, \sigma_u^2)$) and the residuals e_{ij}'s are also independent and identically distributed with mean 0 and variance σ_e^2 (this is, $e_{ij} \sim iidN(0, \sigma_e^2)$).
In addition, e_{ij}'s are independent of u_i's of each other for all i and j (this is, $Cov(u_i, e_{ij}) = 0$).
Note that the normal distribution of u_i's and e_{ij}'s is an assumption. For the estimator in the i^{th} area we get $\overline{Y}_i \approx \overline{X}_i^{'}\beta + u_i + \overline{e}_i$ when area population means $\overline{X}_i$'s are known. Notice that the mean of residuals is added to the random area effects (this is, we could have $u_i + \overline{e}_i = v_i$ where v_i can be taken as the new random area effects).
By assumptions; $E(\overline{e}_i) = 0$.
When both variances (σ_u^2 and σ_e^2) are known, we obtain the Best Linear Unbiased Predictor (BLUP) of the form:

$$\widehat{\overline{Y}}_i = \gamma_i[\overline{y}_i + (\overline{X}_i - \overline{x}_i)^{'}\widehat{\beta}_{GLS}] + (1 - \gamma_i)\overline{X}_i^{'}\widehat{\beta}_{GLS} \tag{25}$$

where $\widehat{\beta}_{GLS}$ is the Generalized Least Squares (GLS) estimator of β and $\gamma_i = \frac{\sigma_u^2}{\sigma_u^2 + n_i^{-1}\sigma_e^2}$.
When the two variances are not known, they are estimated from the sample data. Usually, β is also estimated in the same way. Therefore, we will obtain two estimates whose linear combination (or weighted average) will be more efficient than either of both.
These are the regression synthetic estimate of the form: $\overline{X}_i^{'}\widehat{\beta}$ and sample regression estimate of the form: $\overline{y}_i + \widehat{\beta}(\overline{X}_i - \overline{x}_i)$. Here, the weight γ_i to the sample regression estimate is found/estimated by using the estimates $\widehat{\sigma}_u^2$ and $\widehat{\sigma}_e^2$ of the variances σ_u^2 and σ_e^2 respectively. Then, we get the estimate of the weight defined as follows:
$\widehat{\gamma}_i = \frac{\widehat{\sigma}_u^2}{\widehat{\sigma}_u^2 + n_i^{-1}\widehat{\sigma}_e^2}$ and then $(1 - \widehat{\gamma}_i)$ to the regression synthetic estimate.
That is, the BLUP becomes:
$\widehat{\overline{Y}}_i = \widehat{\gamma}_i[\overline{y}_i + (\overline{X}_i - \overline{x}_i)^{'}\widehat{\beta}_{GLS}] + (1 - \widehat{\gamma}_i)\overline{X}_i^{'}\widehat{\beta}_{GLS}$ [11], [13].
We assume that the parameter of interest is the area mean $\overline{Y}_i$ which will be of the form:

$$\overline{Y}_i = E[\overline{Y}_i|u_i] = \overline{X}_i^{'}\beta + u_i \tag{26}$$

where $\overline{X}_i = \sum_{j=1}^{N_i} \frac{X_{ij}}{N_i}$ is a vector of the area population means of the covariates and having assumed that $\overline{e}_i = \sum_{j=1}^{N_i} \frac{e_{ij}}{N_i} \simeq 0$ for N_i large [16],[13].
The empirical best linear unbiased predictor (EBLUP) of the area mean is given by:

$$\overline{Y}_i = \overline{X}_i^{'}\widehat{\beta} + \widehat{\gamma}_i(\overline{y}_i - \overline{x}_i\widehat{\beta}) \tag{27}$$

where $\widehat{\beta}$ is the EBLUP of β and $\overline{y}_i$ and $\overline{x}_i$ are the sample means for the variable of interest (y_{ij}) and covariates (x_{ij}) respectively. Looking at $\widehat{\gamma}_i$ (which is also obtained by replacing σ_u^2 and σ_e^2 by their respective estimates $\widehat{\sigma}_u^2$ and $\widehat{\sigma}_e^2$), we remark that for large values of n_i, $\widehat{\gamma}_i \to 1$ and hence we can use the sample regression estimate as the area population estimate. In this case, the estimate above can be written as:

$$\overline{Y}_i = \overline{y}_i + (\overline{X}_i - \overline{x}_i)^{'}\widehat{\beta} \tag{28}$$

In the same way, for n_i small, the synthetic regression estimate is needed for the efficiency of the estimate [12].
The MSE of the estimator above in the relation (27) was approximated by Kackar and Harville (1984), Prasad and Rao (1990), and Kenward and Roger (1997) among others and is of the form:

$$MSE(\overline{Y}_i(\delta)) = \sum_{k=1}^{3} g_{ki}(\delta) = g_{1i}(\delta) + g_{2i}(\delta) + g_{3i}(\delta) \tag{29}$$

where $\delta = (\sigma_u^2, \sigma_e^2)^{'}$ [13]. The g_{ki}'s are defined below:

$$g_{1i}(\sigma_u^2, \sigma_e^2) = \gamma_i\sigma_e^2/n_i \tag{30}$$

$$g_{2i}(\sigma_u^2, \sigma_e^2) = (\overline{X}_i - \gamma_i \overline{x}_i)^{'} (\sum_{i=1}^{m} A_i)^{-1} (\overline{X}_i - \gamma_i \overline{x}_i) \tag{31}$$

where $A_i = \frac{1}{\sigma_e^2} \sum_{j=1}^{n_i} (x_{ij} x_{ij}^{'} - \gamma_i n_i \overline{x}_i \overline{x}_i^{'})$

$$g_{3i}(\sigma_u^2, \sigma_e^2) = \frac{\sigma_e^4 \overline{V}_{uu}(\delta) + \sigma_u^4 \overline{V}_{ee}(\delta) - 2\sigma_e^2 \sigma_u^2 \overline{V}_{ue}(\delta)}{n_i^2 (\sigma_u^2 + \sigma_e^2)^3} \tag{32}$$

where $\overline{V}_{uu}$ and $\overline{V}_{ee}$ are the asymptotic variances of σ_u^2 and σ_e^2 respectively and $\overline{V}_{ue}$ is the asymptotic covariance between σ_u^2 and σ_e^2. In practice, $\delta = (\sigma_u^2, \sigma_e^2)$ is replaced by its unbiased estimator $\widehat{\delta} = (\widehat{\sigma}_u^2, \widehat{\sigma}_e^2)$ [15],[13].

3.3.2. Area level model

This type of model is used when only the area level auxiliary information is available. It was first used by Fay and Herriot(1979) for the prediction of mean per capita income (PCI) in small geographical areas. It is of the form:

$$\tilde{\theta}_i = X_i^{'} \beta + u_i + e_i \tag{33}$$

This is called the standard linear mixed model and it can be split into two models: the sampling model also called the matching model and the linking model. The sampling model is as follows:

$$\tilde{\theta}_i = \theta_i + e_i \tag{34}$$

and the linking model is:

$$\theta_i = X_i^{'} \beta + u_i \tag{35}$$

where $X_i = (X_{1i}, X_{2i}, ..., X_{pi})$ is a vector of p area-specific auxiliary variables, $u_i \sim iidN(0, \sigma_u^2)$ and $e_i \sim iidN(0, \sigma_e^2)$. u_i are called model dependent random effects and e_i's are called "sampling errors". $\beta = (\beta_1, \beta_2, ..., \beta_p)$ is the regression parameters vector corresponding to X_i. β can be found by Generalised Least Squares or weighted least squares [12].

From the linear mixed model, we obtain the Best Linear Unbiased Predictor (BLUP) of the following form:

$$\widehat{\theta}_i = \gamma_i \tilde{\theta}_i + (1 - \gamma_i) X_i^{'} \widehat{\beta} = X_i^{'} \widehat{\beta} + \gamma_i (\tilde{\theta} - X_i^{'} \widehat{\beta}) \tag{36}$$

where γ_i is calculated by: $\gamma_i = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$, σ_e^2 is called design variance and σ_u^2 is called model variance. σ_e^2 and σ_u^2 are usually unknown but they are replaced by their sample estimates by use of available data [9],[10].
If there was no sample in the i^{th} area, the estimator $\widehat{\theta}_i = X_i^{'} \widehat{\beta} + \gamma_i (\tilde{\theta} - X_i^{'} \widehat{\beta})$ will be replaced by $\widehat{\theta}_i = X_i^{'} \widehat{\beta}$ which is the regression synthetic estimator [12].

The efficiency of our estimators is tested through the assessment of errors. In this case, we assume the design variance σ_e^2 to be known and we remain with the task of estimating the model variance σ_u^2 and estimation parameter β. It would be better if σ_u^2 and β were also known because the variance of our estimator would be of the form:
$Var(\widehat{\theta}_i) = \gamma_i \sigma_e^2$. Prasard and Rao (1990) approximated the MSE of the predictor when both σ_u^2 and β are replaced by their estimates $\widehat{\sigma}_u^2$ and $\widehat{\beta}$. It was found that:

$$MSE(\widehat{\theta}_i(\widehat{\beta}, \widehat{\sigma}_u^2)) = E[\widehat{\theta}_i(\widehat{\beta}, \widehat{\sigma}_u^2) - \theta_i] = g_{1i} + g_{2i} + g_{3i} \tag{37}$$

with $g_{1i} = \gamma_i \sigma_e^2$, $g_{2i} = (1-\gamma_i)^2 X_i^{'} Var(\widehat{\beta}) X_i$ and $g_{3i} = \frac{\sigma_u^4}{(\sigma_u^2+\sigma_e^2)^3} Var(\widehat{\sigma}_u^2)$. Note that g_{2i} and g_{3i} represent the excess in MSE due due to the estimation of β and σ_u^2 respectively and are of lower order [9].

g_{1i} indicates us that MSE can be reduced through the direct estimate when γ_i is small. In the case of known σ_u^2 and β, we will get much less MSEs since both g_{2i} and g_{3i} are from the estimation of β and σ_u^2 as said above.
In the case of absence of sample in the area where we use the regression synthetic estimatore, the MSE will be the following:

$$MSE(\widehat{\theta}_i) = \widehat{\sigma}_u^2 + X_i \widehat{V}(\widehat{\beta}) X_i^{'} \tag{38}$$

where $\widehat{V}(\widehat{\beta})$ is the variance obtained in the estimation of β.
When estimating the sample variance σ_u^2 which is unknown, some estimates become negative and hence truncated to zero. This will mislead us and have difficulties because of those ignored quantities. The area random effects can be tested to see if they are very meaningful by testing the hypothesis: $H_0 : \sigma_u^2 = 0$ at a specified significance level α. If H_o is not rejected, we can use the synthetic estimator with no random area effects. Otherwise, we will use the model as proposed above [11].
Remark: The estimates which are found are intended to produce: Good ranks, Good histogram and Good area-specific estimates. From these three qualities which are required, the estimates are called:"Triple-goal" estimates.

3.3.3. Mixed logistic model

Now we are going to deal with a binary variable. Previously, we considered a continuous variable (which can take on some values either on a specific interval). When the variable of interest takes only two possible responses (which can be taken as success or failure, present or absent accordingly, 1 or 0), for example when finding areas proportions or totals, we use the mixed logistic model. This type of model was first used by MacGibbon and Tomberlin (1989).
Let us consider the case when we are interested in the i^{th} area proportion

p_i. Then; the area proportion is found as: $p_i = \sum_{j=1}^{N_i} \frac{y_{ij}}{N_i}$. Note that the (binary) variable of interest Y is defined as $P(y_{ij} = 1|p_{ij}) = p_{ij}$ for the j^{th} item in the i^{th} area. In other words, we denote the success probability for the j^{th} item in the i^{th} area by p_{ij}, this is, $p_{ij} = P(y_{ij} = 1)$ where $P(.)$ denotes the probability. We always keep in mind that $y_{ij} = 1, 0$ in this case. Then, the logistic model is written in the following form:

$$logit(p_{ij}) = x_{ij}^{'}\beta + u_i \tag{39}$$

where $u_i \sim iidN(0, \sigma_u^2)$ and $logit(p_{ij}) = log(\frac{p_{ij}}{1-p_{ij}})$. The parameters to be estimated are β and σ_u^2. Once the two parameters are estimated, we get an Empirical Best Predictor (EBP) of the form: $\widehat{p}_i^{EBP} = E[p_i|y_{ij}, \widehat{\beta}, \widehat{\sigma}_u^2]$
Notice that solving the equation $log(\frac{p_{ij}}{1-p_{ij}}) = x_{ij}^{'}\beta + u_i$ for p_{ij} results in: $p_{ij} = \frac{exp(x_{ij}^{'}\beta+u_i)}{1+exp(x_{ij}^{'}\beta+u_i)}$.
Having the estimates of σ_u^2 and β, we can obtain B realizations by Monte Carlo Markov Chain (MCMC) simulations where B is a large number and we operate to get the desired estimator $\widehat{p}_i$ This is, for each $b = 1, 2, ..., B$, we obtain the triplet $(\widehat{\beta}^b, \widehat{\sigma}_u^{2,b}, \{\widehat{u}_i^b\})$ with the corresponding

$$y_{ik} \sim \widehat{p}_{ik} = \frac{exp(x_{ij}^{'}\beta^b + u_i^b)}{1 + exp(x_{ij}^{'}\beta^b + u_i^b)} \tag{40}$$

where $y_{ik} = \frac{\sum_{b=1}^{B} y_{ik}^b}{B}$'s are for the non-sampled elements and hence the area estimate becomes the following:

$$\widehat{p}_i = \frac{\sum_j y_{ij} + \sum_k \widehat{p}_{ik}}{N_i} = \frac{1}{B}\sum_{b=1}^{B}[\sum_j y_{ij} + \sum_k \widehat{p}_{ik}^b]/N_i = \frac{1}{B}\sum_{b=1}^{B}\widehat{p}_i^b \tag{41}$$

where $j \in s_i$ and $k \in (N_i - s_i)$.
After obtaining the residuals (u_i's) distribution through the estimation of their variance σ_u^2, we generated data for the remaining non-sampled part of the area as seen above. Here, $\widehat{y}_{ik}^b$ indicates the b^{th} simulated value of the variable of interest for the k^{th} non-sampled item in the i^{th} area. In general, the superscript b stands for the b^{th} simulation.
As seen in the relations above, the value of Y for the $(N_i - s_i)$ non-sampled items is approximated by its probability to be success. The estimated area proportion $\widehat{p}_i$ is obtained by finding the average of the collected data together with the simulated data. This is why we divided by the total number (N_i) of the items in the area.
The obtained estimate $\widehat{p}_i$ can be trusted due to the fact that it is from a large frame of data provided that the simulated ones were of acceptable precision. This is why we prefer using a large number B of simulations. The approximated posterior variance is $V_{post}(\widehat{p}_i) = \frac{1}{B(B-1)}\sum_{b=1}^{B}(\widehat{p}_i^b - \widehat{p}_i)^2$ [16].

4. Elbers, Lanjouw and Lanjouw (ELL) Method

The Elbers, Lanjouw and Lanjouw (ELL) Method also called the World Bank Method is the statistical technique used in Small Area Estimation used for the population welfare estimators assessment. It was originally proposed by Elbers, Lanjouw and Lanjouw in 2003. In presence of household-level data for the target variable Y, we select a set of covariates to be used to estimate Y. The population is supposed to be divided into non-overlaping clusters. Let Y_{ij} denotes the value of Y for the household j in the cluster i and X_{ij} denotes a set of covariates selected. Preferably, the chosen covariates are common to both survey and the latest census. To estimate the model, we use the log-transformation of Y. That is, we use $\theta_{ij} = log(Y_{ij})$ (or $\theta_{ij} = ln(Y_{ij})$)to obtain the nested error regression model.
Then we obtain the following model:

$$\theta_{ij} = X_{ij}^{'}\beta + u_{ij} \tag{42}$$

where the estimate of β is such that $E(\theta_{ij}|X_{ij}) = X_{ij}^{'}\hat{\beta}$. The error term u_{ij} is decomposed into two terms as $u_{ij} = \eta_i + \varepsilon_{ij}$ with η_i and ε_{ij} representing the cluster-specific effects and household-specific effects respectively. Therefore, the model above is written in the form below:

$$\theta_{ij} = X_{ij}^{'}\beta + \eta_i + \varepsilon_{ij} \tag{43}$$

with $\eta_i \sim iidN(0, \sigma_\eta^2)$, $\varepsilon_{ij} \sim iidN(0, \sigma_{\varepsilon_{ij}}^2)$ and $Cov(\eta, \varepsilon) = 0$ [16].

The two random effects are not only independent to each other, but also uncorrelated to the auxiliary variables X_{ij}. In estimating the variance σ_{ij}^2 for the random effects $u_{ij} = \eta_i + \varepsilon_{ij}$, the greater the fraction due to the common component η_i the less one enjoys the benefits of aggregating over more households within a cluster [16].
The estimates of the two components of u_{ij} are found as follows:
$\widehat{u}_{ij} = \widehat{u}_{i.} + (\widehat{u}_{ij} - \widehat{u}_{i.}) = \widehat{\eta}_i + \widehat{\varepsilon}_{ij}$
The variance for ε_{ij} is estimated by a logistic form below:

$$\sigma_\varepsilon^2 = \frac{Ae^{{z_{ij}}'\alpha} + B}{1 + e^{{z_{ij}}'\alpha}} \tag{44}$$

[10].
We choose using the logistic form to avoid negative values for the variance or its extremely high predicted values. The constants A and B are upper and lower bounds to avoid extreme values for the variance and z_{ij} is a set of vector of household characteristics which is not necessarily different from x_{ij}.

It was found that the values $B = 0$ and $A = 1.05 \max(\sigma^2_{\varepsilon_{ij}})$ give good estimates of parameters. We remark that the parameter α is to be estimated. Using these values given above, the task becomes easier than when we were to estimate both the parameter α and constants A and B.
We need to find the residuals terms η and ε by simulations. Introducing the standardized household residuals:

$$e^*_{ij} = \frac{\varepsilon_{ij}}{\widehat{\sigma}_{\varepsilon_{ij}}} - (1/H) \sum_{i,j} \frac{\varepsilon_{ij}}{\widehat{\sigma}_{\varepsilon_{ij}}} \tag{45}$$

we find the appropriate distributional forms.
H represents the number of observations [16],[14].
β in the model (42)is estimated by Ordinary Least Squares or Weighted Least Squares estimation. On the other hand, using the model (42), β, σ^2_η and $\sigma^2_{\varepsilon_{ij}}$ are obtained using the Restricted Maximum Likelihood (REML)estimation. When available, the three estimates will allow us to generate a bootstrap populations, let's say K populations in the following way:

$$\theta^{*,k}_{ij} = X^{'}_{ij}\beta + \eta^*_i + \varepsilon^*_{ij} \tag{46}$$

for $k = 1, 2, ...K$ [19]. Note that the η^*_i's and ε^*_{ij}'s are obtained from the distributions $N(0, \hat{\sigma_\eta}^2)$ and $N(0, \hat{\sigma_\varepsilon}^2)$ respectively.
For the i^{th} cluster, we calculate the corresponding estimate θ^*_i from the $\theta^{*,k}_{ij}$'s and then we obtain the following:

$$\hat{\theta}_i = K^{-1} \sum_{k=1}^{K} \theta^{*,k}_i \tag{47}$$

With MSE found as below:

$$MSE(\hat{\theta}_i) = K^{-1} \sum_{k=1}^{K} (\theta^{*,k}_i - \hat{\theta}_i)^2 \tag{48}$$

[9].
In the presence of both the household-level covariates X_{ij} and location-level covariates Z_i, the model (42) can be written as:

$$\theta_{ij} = X^{'}_{ij}\beta + Z_i\gamma + u_{ij} \tag{49}$$

where both β and γ are vectors of the associated regression coefficients respectively. The estimates of these coefficients are such that $E(\theta_{ij}|X_{ij}) = X^{'}_{ij}\hat{\beta} + Z_i\hat{\gamma}$.

Anthropometric model:

This has been studied by Fujii et al.(2004)[20]. Here, we are going to deal with anthropometric measures. These are the measures related to human body as the name indicates (anthropo:human, metric:measure). As we are interested in Stunting, Wasting and Underweight; we will take under consideration the height and weight of children under five years old.
We denote our variable of interest by Y where we write Y_{chi} to represent the measure of the i^{th} individual (child) in the h^{th} household in the c^{th} cluster. Since the notation will be the same for all the variables of interest, we put the superscript l and get $Y_{chi}^{(l)}$ to mean the variables separately where $l = 1, ..., L$ and L is the number of all variables of interest.
The anthropometric model is written in the form below:

$$Y_{chi}^{(l)} = X_{chi}^{(l)}\beta + u_{chi}^{(l)} \tag{50}$$

where $u_{chi}^{(l)}$ is the residual term which is such that $E[u_{chi}^{(l)}|X_{chi}^{(l)}] = 0$. In addition, $u_{chi}^{(l)}$ can be split into three components where we have the cluster-specific effects $\eta_c^{(l)}$, household-specific effects $\epsilon_{ch}^{(l)}$ and individual-specific effects $\delta_{chi}^{(l)}$.
Then, we have:

$$u_{chi}^{(l)} = \eta_c^{(l)} + \epsilon_{ch}^{(l)} + \delta_{chi}^{(l)} \tag{51}$$

where $\eta_c^{(l)}$, $\epsilon_{ch}^{(l)}$ and $\delta_{chi}^{(l)}$ are all assumed to be random variables and are such that:

$$E[\eta_c^{(l)}] = E[\epsilon_{ch}^{(l)}] = E[\delta_{chi}^{(l)}] = E[u_{chi}^{(l)}] = 0$$
$$Cov(\eta_c^{(l)}, \epsilon_{ch}^{(l)}) = Cov(\eta_c^{(l)}, \delta_{chi}^{(l)}) = Cov(\epsilon_{ch}^{(l)}, \delta_{chi}^{(l)}) = 0$$

where the last equation shows that the three components of the residuals are not correlated. We work under the following assumptions on the variance matrices of these effects:
$Var(\eta_c)$ and $Var(\epsilon_{ch})$ are diagonal while $Var(\delta_{chi})$ may not. This is logically clear. We first remark that η_c, ϵ_{ch} and δ_{chi} represent the vector of their components in all the L variables. In other words, they can be written as:

$$\eta_c = (\eta_c^{(1)}, \eta_c^{(2)}, ..., \eta_c^{(L)})'$$
$$\epsilon_{ch} = (\epsilon_{ch}^{(1)}, \epsilon_{ch}^{(2)}, ..., \epsilon_{ch}^{(L)})'$$
$$\delta_{chi} = (\delta_{chi}^{(1)}, \delta_{chi}^{(2)}, ..., \delta_{chi}^{(L)})'$$

Hence, it is understandable that the covariance (which is the source of correlation) is considered in variables for the same individual and no need of it at the household and cluster levels. Then, we have: $E[\eta_c^{(l)}, \eta_c^{(m)}] = E[\epsilon_{ch}^{(l)}, \epsilon_{ch}^{(m)}] = 0 \ \forall l \neq m$. This is why we can write the two covariance matrices as:

$$Var(\eta_c) = diag((\sigma_\eta^l)^2)$$
$$Var(\epsilon_{ch}) = diag((\sigma_\epsilon^l)^2)$$

As it is general for any covariance matrix, $Var(\delta_{chi})$ is symmetric. In most case, we use the notation: $Var(\delta_{chi}) = \sigma_\delta^{(l,m)}$ to show the covariance of the l^{th} and the m^{th} variables and $\sigma_\delta^{(l,l)} = (\sigma_\delta^l)^2$ to show the variance for the l^{th} variable for the same individual. Hence, we write the variance-covariance matrix as: $Var(\delta_{chi}) = (\sigma_\delta^{(l,m)})$ for all l, $m \in \{1,...,L\}$.
Let us denote by C the number of all clusters where each cluster $c(1 \leqslant c \leqslant C)$ is made of H_c households. The number of individuals in the h^{th} household in the c^{th} cluster is denoted by I_{ch}. Therefore, the total number of observations will be: $N = \sum_{c=1}^{C} \sum_{h=1}^{H_c} I_{ch}$. Each cluster has its own weight w_c and all weights are such that they sum up to 1 (this is, $\sum_{c=1}^{C} w_c = 1$).
The simple means of these three components or residuals are found as usual.

$$u_{ch.} = \frac{1}{I_{ch}} \sum_{i}^{I_{ch}} u_{chi} = \frac{1}{I_{ch}} \sum_{i=1}^{I_{ch}} (\eta_c + \epsilon_{ch} + \delta_{chi}) = \eta_c + \epsilon_{ch} + \delta_{ch.}$$
$$u_{c..} = \frac{1}{H_c} \sum_{h=1}^{H_c} u_{ch.} = \sum_{h=1}^{H_c} (\eta_c + \epsilon_{ch} + \delta_{ch.}) = \eta_c + \epsilon_{c.} + \delta_{c..},$$

where

$$\delta_{ch.} = \frac{1}{I_{ch}} \sum_{i}^{I_{ch}} \delta_{chi}$$
$$\epsilon_{c.} = \frac{1}{H_c} \sum_{h=1}^{H_c} \epsilon_{ch}$$
$$\delta_{c..} = \frac{1}{H_c} \sum_{h=1}^{H_c} \delta_{ch.}$$

Variances of the three components of residuals are found as below:

$$\sigma_\delta^2 = E\big[\sum_c \frac{w_c}{H_c} \sum_h \sum_i \frac{(u_{chi} - u_{ch.})^2}{I_{ch} - 1}\big],$$
$$\sigma_{\epsilon,ch}^2 = \frac{H_c.E\big[(u_{ch.} - u_{c..})^2\big]}{H_c - 2} - \frac{E\big[\sum_{h'}(u_{ch'.} - u_{c..})^2\big]}{(H_c - 2)(H_c - 1)} - \frac{\sigma_\delta^2}{I_{ch}},$$
$$\sigma_\eta^2 = \frac{\sum_c w_c H_c E\big[(u_{c..})^2\big] - \sum_c \frac{w_c}{H_c} \sum_h E\big[(u_{ch.})^2\big]}{\sum_c w_c(H_c - 1)}.$$

The first equation is valid for the household with $I_{ch} > 1$. The second and third equations are valid for households with $I_{ch} > 2$.
For two anthropometric variables l and m, we can find the intra-personal correlation denoted by $\sigma_\delta^{l,m}$ as follows:

$$\sigma_\delta^{(l,m)} = \sum_c \frac{w_c}{H_c} \sum_h \sum_i \frac{E\big[(u_{chi}^{(l)} - u_{ch.}^{(l)})(u_{chi}^{(m)} - u_{ch.}^{(m)})\big]}{I_{ch} - 1}$$

or equivalently,

$$\sigma_\delta^{(l,m)} = \sum_c \frac{w_c}{H_c} \sum_h \sum_i \frac{E[u_{chi}^{(l)}.u_{chi}^{(m)}]}{I_{ch}}.$$

We apply the Ordinary Least Squares (OLS) method to find the regression coefficient β for each variable/ indicator. This is, for the k^{th} indicator, we had: $Y_{chi}^{(l)} = X_{chi}^{(l)}\beta^{(l)} + u_{chi}^{(l)}$ and after getting the corresponding regression coefficient $\beta_{OLS}^{(l)}$, we estimate the residuals as follows: $\widehat{u}_{chi}^{(l)} = y_{chi}^{(l)} - X_{chi}^{(l)}\widehat{\beta}^{(l)}$. From this, we define $\widehat{u}_{ch.}^{(l)}$ and $\widehat{u}_{c..}^{(l)}$ as seen previously where u is now replaced by $\widehat{u}$.
It may happen that the variance for the cluster effects is negative. In this case, we drop it to zero and consequently, we fall in the case where the cluster (location)-specific effects are not considered (are assumed to be zero). We remain only with the household effects and individual effects. The residual term becomes: $u_{chi}^{(l)} = \epsilon_{ch}^{(l)} + \delta_{chi}^{(l)}$.
It sometimes arrive that even these two effects (household effects and individual effects) are not different. This is in the case where the number of individuals in a household is equal to one ($I_{ch} = 1$). In this last case combined with the previous one (in absence of cluster effects), we will remain with the mixture (sum) of the household effects and individual effects expressed as follows: $s_{\epsilon,ch}^2 = \sigma_{\epsilon,ch}^2 + \sigma_\delta^2$.

Having replaced $u_{chi}^{(l)}$ by its estimate $\widehat{u}_{chi}^{(l)}$, we get the estimates of all the effects as defined above. The estimate of our new expression of household and individual effects will be of the form below:

$$\widehat{s}^2_{\epsilon,ch} = \frac{H_c.E\big[(\widehat{u}_{ch.} - \widehat{u}_{c..})^2}{H_c - 2} - \frac{E\big[\sum_{h'}(\widehat{u}_{ch'.} - \widehat{u}_{c..})^2\big]}{(H_c-2)(H_c-1)} + \frac{I_{ch}-1}{I_{ch}}\widehat{\sigma}^2_{\delta} \qquad (52)$$

where we had: $\widehat{s}^2_{\epsilon,ch} = \widehat{\sigma}^2_{\epsilon,ch} + \widehat{\sigma}^2_{\delta}$. Note that $\widehat{\sigma}^2_{\epsilon,ch}$ and $\widehat{\sigma}^2_{\delta}$ are the estimates of $\sigma^2_{\epsilon,ch}$ and σ^2_{δ} respectively as said above.

In absence of cluster effects, we have: $E[(u_{ch.})^2] = \sigma^2_{\epsilon,ch} + \frac{\sigma^2_{\delta}}{I_{ch}}$.
Hence, we get the following expression for the household and individual effects:

$$\widehat{s}^2_{\epsilon,ch} = \widehat{u}^2_{ch.} + \frac{I_{ch}-1}{I_{ch}}\widehat{\sigma}^2_{\delta} \qquad (53)$$

We then use the logistic equation (44) written as:

$$s^2_{\epsilon,ch} = \sigma^2_{\varepsilon_{ij}} = \frac{Ae^{{z_{ij}}'\alpha} + B}{1 + e^{{z_{ij}}'\alpha}}. \qquad (54)$$

Distribution of residual components:

For the indicator under consideration, we need the distribution of its residual terms. For this, we first define: $\widehat{e}^2_{ch} = \frac{H_c-2}{H_c}(\widehat{\sigma}^2_{\epsilon,ch} + \frac{\widehat{\sigma}^2_{\delta}}{I_{ch}}) + \frac{1}{H_c^2}\sum_{h'}(\widehat{\sigma}^2_{\epsilon,ch'} + \frac{\widehat{\sigma}^2_{\delta}}{I_{ch'}})$ which is considered as the estimate of $E[(u_{ch.} - u_{c..})^2]$.
For families with at least one child under five years (because this is our target population), we have:

$$E[u^2_{c..}] = \sigma^2_{\eta} + \frac{1}{H_c(H_c-1)}E[\sum_{h}(u_{ch.} - u_{c..})^2].$$

Defining:

$$\widehat{Var}(\widehat{u}_{c..}) = \widehat{\sigma}^2_{\eta} + \frac{1}{H_c(H_c-1)}\sum_{h}\widehat{e}^2_{ch}$$

and

$$w^*_c = w_c\frac{\sum_c H_c}{\sum_{c^*} H_{c^*}}$$

where the summation over c^* indicates the summation done having considered only clusters with more than one household ($H_c > 1$).

We get the following distributions:

$$\tilde{\eta}_c = \frac{\widehat{u}_{c..}}{\sqrt{\widehat{Var}(\widehat{u}_{c..})}} - \sum_{c'} \frac{w^*_{c'}.\widehat{u}_{c'..}}{\sqrt{\widehat{Var}(\widehat{u}_{c..})}}$$

$$\tilde{\epsilon}_{ch} = \frac{\widehat{u}_{ch.} - \widehat{u}_{c..}}{\sqrt{\widehat{e}^2_{ch}}} - \sum_{c'} \frac{w^*_{c'}}{H_{c'}} \sum_{h'} \frac{\widehat{u}_{c'h'.} - \widehat{u}_{c'..}}{\sqrt{\widehat{e}^2_{c'h'}}}$$

$$\tilde{\delta}_{chi} = \frac{\widehat{u}_{chi} - \widehat{u}_{ch.}}{\sqrt{\frac{I_{ch}-1}{I_{ch}}\widehat{\sigma}^2_\delta}}$$

The particular case when there are no cluster effects is also investigated. For the unmodeled cluster effects, we get:

$$var(u_{ch.}) = E[(u_{ch.})^2] = \sigma^2_{\epsilon,ch} + \frac{\sigma^2_\delta}{I_{ch}}$$

and errors distributions become:

$$\tilde{\epsilon}_{ch} = \frac{\widehat{u}_{ch.}}{\widehat{\sigma}^2_{\epsilon,ch} + \frac{\widehat{\sigma}^2_\delta}{I_{ch}}} - \sum_{c'} \frac{w_{c'}}{H_{c'}} \sum_{h'} \frac{\widehat{u}_{c'h'.}}{\sqrt{\widehat{\sigma}^2_{\epsilon,ch} + \frac{\widehat{\sigma}^2_\delta}{I_{ch}}}}$$

$$\tilde{\delta}_{chi} = \frac{\widehat{u}_{chi} - \widehat{u}_{ch.}}{\sqrt{\frac{I_{ch}-1}{I_{ch}}\widehat{\sigma}^2_\delta}}$$

[20].

5. Empirical Data Analysis

5.1. Data description

In this thesis, we were aiming at finding the estimates of malnutrition indicators which are: stunting, wasting and underweight at district level. The districts were taken as small areas. This is, we targeted at finding the estimates at district level of those three indicators. Unfortunately, we have not been able to conduct a survey in order to get updated information. We have used the DHS 2010 data as our present data because we used them in the model and linked them to the covariates which were found appropriate. In this survey, they had collected the heights and weights of children under five years old with their respective ages. They had calculated the z-scores for the three quantities of interest and this made the task easier. The z-scores were for height_for_age, weight_for_height and weight_for_age.
Those z-scores helped us to find the direct proportions of their corresponding indicators. These direct proportions are then considered as present (Y_i^l's).
The data on poverty rates used were from the NISR report entitled: Rwanda poverty profile report 2013/2014 which is the results of the fourth integrated households living conditions survey (EICV 4 in French abbreviations).
Recent proportions of these malnutrition indicators were taken from the 2010 Rwanda Demographic and Health Survey in the report published in February 2012. The literacy rates were found in the 2014 Education Statistical Yearbook published by Ministry of Education in March 2015.
The urbanization rates were found in the integrated district development planning: Situation Analysis for Karongi District whose final version was published in December, 2014.
It was found that malnutrition is influenced by mothers level of education. This variable would be used as a covariate to explain the level of malnutrition in a given district. For this, we used the illiteracy rate in each district and relate it to the malnutrition rate. In the model, we used the literacy rate and the information is the same since if r % are literate, then (100-r) % are illiterate. This is why the corresponding coefficient is negative apart from for the third indicator (this is critical). It is not defendable because the direct proportions were zero in some districts and the estimates were from the indirect estimates. Moreover, from the fact that the literacy rate is shown as a factor of underweight since the corresponding coefficient is positive, we can desire having much information. However, we can accept that the proportions of underweight are below compared to those of stunting and wasting.
The household living conditions were also found to be the major factor of malnutrition. For this, we used poverty rate in a district to relate it to the malnutrition rate.

5.2. Data analysis

As said previously, we have to use the ELL method to conduct the analysis in order to find the estimates. We will use the covariates whose information is known at district level.
As we have area-level covariates; we have the area-level model (as expressed by the equation (35))where the proportions of each indicator (at district-level) are linked to those of the covariates. Considering the model (42), we obtain the area-level model $\theta_i = X_i^{'}\beta + u_i$ where X_i is the set of three covariates found to be strongly linked to malnutrition. Note that the vector of coefficients β is found by Least squares errors method. For each indicator, we separately found the corresponding model even if the covariates were the same. With the data set, we found the direct estimates proportions of malnourished children by finding those whose z-scores are below -2SD. Note that we did not differentiate the severe and mild malnutrition. This is why we did not separate children whose z-scores are below -3SD (severely malnourished children) from those whose z-scores are below -2SD (mildly malnourished children). We only focused on malnutrition in general independently of its level.

Summary table of the obtained results: Estimates of Malnutrition indicators at district level compared to those obtained in 2010.

	Stunting	Underweight	Wasting	Stunt(10)	Underw(10)	Wast(10)
Nyarugenge	15.3	6.6	2.4	28.3	5.7	3.5
Gasabo	17.3	8.9	3.1	23.8	10.8	6.2
Kicukiro	9.7	3.1	1.4	18.9	3.9	2.4
Nyanza	36.2	16.9	1.2	47.6	17.8	1.5
Gisagara	43.0	19.2	3.8	45.4	13.7	8.5
Nyaruguru	41.9	18.43	1.42	45	11.6	1
Huye	32.9	15.4	0.4	53.5	19.7	2.3
Nyamagabe	38.9	17.4	0.4	20.7	8.1	5.1
Ruhango	34.9	17.2	3.3	46.7	7.7	6.5
Muhanga	31.5	15.8	1.7	45.3	10.9	1.4
Kamonyi	30.1	15.8	1.4	56.7	13.1	2.7
Karongi	39.3	18.2	3.0	60.3	16.4	5.6
Rutsiro	42.0	19.3	4.3	54.9	10.3	2.6
Rubavu	32.3	12.6	0.0	51.5	9.9	0.7
Nyabihu	36.2	17.0	2.5	53.4	14.3	0.7
Ngororero	41.8	18.8	2.8	53.4	14.3	1.8
Rusizi	34.1	16.9	2.4	40.9	14.4	2.7
Nyamasheke	44.0	21.7	10.4	33.2	9.4	5.6
Rulindo	38.9	19.4	6.6	42.9	15.3	2.6
Gakenke	36.7	18.4	4.9	63.6	12.4	0.7
Musanze	32.3	14.8	2.3	45.3	7	0.7
Burera	41.9	18.9	3.4	52	10.5	0.9
Gicumbi	38.6	16.2	5.1	46.6	8.8	1.4
Rwamagana	29.7	15.0	0.3	29.2	5.4	5.7
Nyagatare	37.9	17.7	3.5	42.2	8.0	0.9
Gatsibo	36.3	15.7	2.3	51.5	10.6	2.6
Kayonza	30.9	15.0	0.0	44.5	15.6	4.1
Kirehe	36.3	16.4	2.0	50.7	13.2	1.5
Ngoma	39.2	18.0	3.8	50.2	15.8	4.2
Bugesera	33.7	16.2	1.3	38.0	13.4	4.4

where Stunt refers to stunting, Underw to Underweight and Wast to Wasting. The index 10 indicates that the results are from DHS(2010).

6. Conclusion and recommendations

6.1. Conclusion

There is inequality in districts of Rwanda. This is to mean that all districts are not likely equally poor. In case of using funds for aid, it is not wise to think on countrywide level but focus on some more vulnerable/ suffering districts. Malnutrition is found to be of long time and this shows how much attention is needed.
Among all the provinces, the Kigali City is less suffering from malnutrition as it is seen on the estimated proportions of each of its three districts.
As it was found in other previous researches, the same remark is that the great part of malnutrition observed at high level/ proportion is the chronic one referring to stunting proportions obtained in all districts.

6.2. Recommendations

• The malnutrition of a short recent period is easily recovered once conditions of life become favourable. This is why the 3% of wasted and 16% of people under-weighted are not terrifying but serious measures and precautions are needed to eradicate/minimize the proportion of stunting (35% average at national level) which is still substantial.

• The chronic malnutrition (explained by stunting) being not cured, it can be advised to policy-makers to prevent it in the future generation. This requires much attention because it is at high proportion in all districts compared to the two other remaining indicators.

• In funds sharing if any, the focus would be to those districts with proportions higher than the national average proportion. Among those districts, the six first ones are: Nyamasheke, Gisagara, Rutsiro, Nyaruguru, Burera and Ngororero.

• The estimates obtained can be used in decision-making but we recommend that in the future researches, they may conduct an up-to-date survey to have much more precise, accurate and up-to-date estimates. For this, the variables such as: the household income, parents level of education, habitation region (urban or rural), etc must be included to get more reliable estimates.

• There is still need of improvement in nutrition policy taking into account the fact that the chronic malnutrition is at a remarkable proportion in the whole country. Whatever cost of hunger, there must be funds for child malnutrition eradication or reduction having known that the nutritional status affects much the child performance in different activities. This is from the fact that the chronic malnutrition is the one observed at high proportion.

• It will be much more helpful to find sector level estimates since the district is so wide that it can have sectors whose life conditions are different. It can be tested if the difference in life conditions of sectors of a specific district are significantly different. If so, the recommended estimates may be more informative and helpful than those of district level.

References

[1] NISR, MOH (2015). *Rwanda Demographic and Health Survey,2014-15. Key indicators*

[2] MINAGRI et al (2014). Cost of hunger Study in Rwanda: Child Undernutrition in Rwanda Implications in Achieving Vision 2020. *Third National Food and Nutrition Summit 2014*

[3] USAID (2014). Rwanda: Nutrition Profile

[4] MINALOC, MOH, MINAGRI (2013). *Rwanda National Food and Nutrition Policy. Post Validation and Adjustment DRAFT*

[5] NISR (2015). *Rwanda Poverty Profile Report. Integrated Household Living Conditions Survey (EICV)*

[6] UNICEF (2013). *Improving Child Nutrition. The achievable imperative for global progress*

[7] MINAGRI et al.(2013). *Comprehensive Africa Agriculture Development Programme (CAADP). Nutrition Country Paper Paper-Rwanda(Draft).*(2013)

[8] Ministry of Health (2011). *The Implementation of In-Home Fortification and Nutrition Education to Combat Anaemia and Micronutrient Deficiencies Among Children 6-23 Months in Rwanda Phase 1 Final Report.*(2011)

[9] Rao J. N.K.(2003). *Small area estimation.* Carleton University.

[10] Chris Elbers, Jean O. Lanjouw, Peter Lanjouw.*Micro-level estimation of poverty and inequality*

[11] Prof. Monica Pratesi (2012). *Recent Development in Small Area Estimation Methodology.* Department of Statistics and Mathematics Applied to Economics, University of Pisa. Valmiera, 2012.

[12] Azizur Rahman (2008). *A Review of Small Area Estimation Problems and Methodological Developments.* University of Canberra.

[13] Hukum Chandra. *Overview of small area estimation Techniques.* Indian Agricultural Statistics Research Institute, New Delhi-110012.

[14] J.N.K.Rao (2012). *Small Area Estimation: Methods and Applications.* Carleton Univerity,Ottawa, Canada.

[15] Pushpal K Mukhopadhyay and Allen McDowell (2011). *Small Area Estimation for Survey Data Analysis Using SAS Software*. SAS Institute Inc., Cary, NC (2011).

[16] Danny Pfeffermann (2013). *New Important Developments in Small Area Estimation*.

[17] Arman Bidarbakht Nia (2011). *Examining poverty and inequality across small areas of islamic republic of iran; tool-making for efficient welfare policies*.

[18] Peter Lanjouw et al. (2004). *Micro-level Estimation of Prevalence of Child Malnutrition In Cambodia*.

[19] Sumonkanti Das. (2013). *An Overview of Poverty Mapping with Application to Bangladesh*. MPE 2013: The Conference(July 09,2013).

[20] Fujii et al.(2004). *Micro-level Estimation of Prevalence of Child Malnutrition In Cambodia*. University of California, World Bank and ORC Macro

[21] M. Ghosh and J. N. K. Rao (1994). *Small Area Estimation: an appraisal*. Statistical Science 1994, Vol. 9, No, 1