Integrated Agricultural and Household Surveys: The Experience of Ethiopia

[1]Biratu Yigezu and [2]Samia Zekaria

Central Statistical Agency of Ethiopia

Addis Ababa

In order to provide reliable and timely data on agriculture, implementing a statistically sound methodology is crucial. This needs investigating the available scientific methods recommended globally and materializing according to the context of the country. Consequently, the Central Statistical Agency (CSA) of Ethiopia has been generating data on agriculture since the early 1980's using the recommended methodology.

The government of Ethiopia has allocated a progressively significant budget for implementing the statistical plans. This was operationally possible due to the establishment of a properly functioning statistical system in the country, under the National Integrated Household Survey Program (NIHSP) by the CSA.

The NIHSP consists of a series of multi-subject surveys, being undertaken on successive occasions. The major contents/subjects covered are: agriculture; labour force; demographic characteristics; household income, consumption and expenditure; welfare monitoring; household based enterprise surveys; farm-gate prices; etc.

The overall sample strategy was carefully designed for the program to facilitate the integration and linking of different surveys conducted in the same period (year), collect data on each survey from the same set of sample of primary sampling units (PSUs) and households, adopt the same concepts, definitions and classification systems and incorporate core items in each of the schedules prepared for the conduct of various surveys. The sampling frame used for the selection of PSU is compiled from the national population and housing census.

The agricultural statistics data produced by the CSA fulfils the vision of global strategy. Moreover, the Master Sample Frame that of NIHSP forms the foundation for the integrated survey framework. The framework also considers the need for the linkage between different data items not only between themselves.

In Ethiopia the quality of statistical data including data on agriculture is measured using the internationally recommended eight dimensions. CSA also implementing and integrating several advanced technologies, like, area frame, multiple frame, small area estimation methodologies and improved stratification.

This paper describes the historical background and the context of global strategy to improve agricultural statistics in Ethiopia, integration of agricultural statistics into the national statistical system, implementation of a master sample frame, quality assurance of agricultural statistics and advanced methodologies planned to be implemented in country.

Key Words: NIHSP, Mater Sample, Agriculture, Multi-subject surveys, Quality, Advanced Methods

[1] Director, Agriculture and Natural Resources Statistics Directorate, e.mail: ybiratu@yahoo.com
[2] Director General, Central Statistical Agency of Ethiopia, e.mail: samiaz@ethionet.et

Acronyms

CSA	Central Statistical Agency of Ethiopia
MoA	Ministry of Agriculture
MoARD	Ministry of Agriculture and Rural Development
CSO	Central Statistical Office
PSU	Primary Sampling Unit
SSU	Secondary Sampling Unit
MTNSP	Medium Term National Statistical Program
NSDS	National Strategy for Development of Statistics
EA	Enumeration Area
RIHSP	Rural Integrated Household Survey Program
NIHSP	National Integrated Household Survey Program
IT	Information Technology
ENADA	Ethiopian National Data Archive
GDP	Gross Domestic Product
FAO	Food and Agriculture Organization
AF	Area Frame
MF	Multiple Frame
PPS	Probability Proportional to Size
GIS	Global Information System
EC	European Commission
RICS	Rural Investment Climate Survey
LSMS – ISA	Living Standards Measurement Study-Integrated Surveys on Agriculture
WB	World Bank

1. Introduction

Geographically, Ethiopia is relatively large country with a total land area of 1.127 million square kilometer. The population of Ethiopia is estimated at around 80 million, giving a population density of nearly 71 persons per square kilometer. As in many other developing countries, the population is predominantly rural, and engaged in agriculture. Only about 16 percent of the populations live in urban areas.

Administratively, the country is divided into nine regional states and two city administrations. Each regional state is divided into administrative zones and zones are further divided into woredas (districts). Under woreda there are different number of kebeles that forming the lowest administrative units of the country.

Agriculture is the main sector of the Ethiopian economy. It accounts for approximately 45 percent of the gross domestic product (GDP), provides employment for 80 percent of the population and generates about 90 percent of the export earnings. Crop production is estimated to contribute on average 60 percent, livestock accounts for 27 percent and forestry and other sub-sectors account for 13 percent of the total agricultural value. This means that any accurate analyses of different elements that influence food security need to be based on reliable indicators from the agricultural sector.

In order to provide reliable and timely data on agriculture, implementing a statistically sound methodology is crucial. This needs investigating the available scientific methods recommended globally and materializing according to the context of the country. Consequently, the Central Statistical Agency (CSA) of Ethiopia has been generating data on agriculture since the early 1980's using the FAO recommended methodology. This exercise was being strengthened through adequate revisions and updating. The development of science and technology has contributed a lot for the development of agricultural statistics. This new methodologies have to be tested well and implemented accordingly. Moreover, the CSA is in the process of implementing various new methods which are assumed to simplify and improve the agricultural statistics in the country.

Given the limited resource that the country can invest for statistics and considering the importance of generating socio-economic and demographic data for development programs, Ethiopia has been implementing the National Integrated Household Survey Program since 1980's.

This first section of this paper reviews briefly the historical background of agricultural statistics in Ethiopia over time presenting progressive improvements in producing data related to food availability. The second section examines the status of agricultural statistics of the CSA in the context of global strategy to improve agricultural statistics whether it is following a standard and acceptable procedure in providing reliable data to the policy makers. The data quality issues on agricultural statistics generated by the CSA are discussed in the third section. Information related to the efforts made at the CSA in implementing advanced technology to improve the timeliness and accuracy is given in the fourth section and concluding remarks in last section.

2. Historical Background of Agricultural Statistics in Ethiopia

A brief history of agricultural statistical programs in Ethiopia illustrates how improvements are continually being made based on technological advancements. Four distinct periods of agricultural statistical developments can be characterized as follows:

a) Prior to 1974: "Ad-hoc surveys"
b) 1974 – 1979: "National (Agricultural) Sample Surveys"
c) 1980 – 1992: "Rural Integrated Household Survey Program," "Integrated System of Food and Agricultural Statistics Program,"
d) 1992 – 2010: "National Integrated Household Survey Program"
e) 2008 and on: Introduction of Area Frame sampling on pilot basis with advanced technologies.

Prior to 1974: Statistical services in general and agricultural statistics in particular were not adequate with respect to coverage, timeliness and reliability of data. This was a period in which agricultural statistics were just beginning to be incorporated into programs. FAO provided consultants and both the then Ministry of Agriculture (MoA) and Central Statistical Office (CSO) developed methodology.

Between 1974 to 1979: The MoA continued to obtain assistance from FAO and conducted six Annual Agricultural Sample Surveys that included a 1976/77 small-scale agricultural sample census. Regional and field supervisors of the MoA and temporary enumerators undertook the fieldwork during each survey. The CSO cooperated closely with the MoA by supplying supervisors, vehicles and equipment. MoA staff members processed the collected data manually.

Between 1980 to1992 and 1992 onwards: Before 1980, available socioeconomic and demographic data in Ethiopia were seriously deficient and out of date. There was no national statistical program to ensure a continuous flow of socio-economic and demographic data needed for planning, monitoring and evaluation of on-going development programs such as poverty reduction. As a result, under Rural Integrated Household Survey Program (RIHSP) the CSA (the then CSO) carried-out 13 Annual Agricultural Sample Surveys between the years 1980/81 to

1992. On the other hand, the National Integrated Household Survey Program (NIHSP) that was established from 1992 onwards enabled the CSA to run a number of annual socio-economic and demographic surveys**.** As a result, the CSA could be able to fully utilize its available infrastructure, field staff (enumerators, supervisors, and drivers), logistics support, and data processing facilities very efficiently and cost effectively.

The NIHSP consists of a series of multi-subject surveys, being undertaken on successive occasions. According to the priorities set up, the major contents/subject covered include: Agriculture; Labour Force; Household Income, Consumption and Expenditure; Welfare Monitoring, and Prices. These surveys are covered in different periods. Agricultural sample surveys are conducted on annual basis and others are surveyed every two, three and five years. The integrated household surveys cover the entire country except some areas inhabited by nomadic population which require different methodological approach.

3. Agricultural Statistics of Ethiopia in the Context of Global Strategy to Improve Agricultural Statistics

Under the auspices of the United Nations Statistical Commission, global strategy to improve agricultural statistics has been drafted due to decline in the quantity and quality of data coming from national statistical offices and/or Ministries of Agriculture. According to the World Development Report, three-fourths of the poor people in developing countries live in rural areas and most depend on agriculture for their livelihood. The importance of agriculture in the effort to reduce poverty places agriculture at the centre of the development agenda. While agriculture is the primary source to feed, clothe, and provide materials for fuel and housing for a growing world population, the challenge is at the same time to lift millions of people out of poverty and hunger, reduce the impact of agriculture on the environment and global warming, and sustain water and land resources. These are issues that go beyond national boundaries.

This increases the need for monitoring and evaluation tools to learn what does and does not work. These tools require much of the basic data countries should be providing any way, but are often not available. Hence, it is important to assess the status of Ethiopian Agricultural Statistics on the context of global strategy that aims to improve the quality and quantity of relevant indicators both nationally and internationally.

3.1 Integrating Agricultural Statistics into the National Statistical System

The purpose of the global strategy to improve agricultural statistics is to provide the vision for national and international statistical systems to produce the basic data and information to guide the decision making required for the 21st century. This strategic plan stands on three pillars, the selection of an agreed upon set of indicators, their integration into the national statistical system, and the methodology to measure them.

One could examine the status of Ethiopia's Agricultural Statistics in the context of global strategy. According to the steps to implement the draft strategic plan it will depend upon the statistical capacity of each country. Hence countries have been grouped in three:

- Those needing to reform their statistical system will begin with the core items and build the rest over time.
- The next group is the countries with National Strategies for the Development of Statistics being implemented. These national strategies need to be reviewed in light of the strategic plan and revised if necessary.
- The third group includes the countries with developed statistical systems.

However, many of these do not meet the integration requirement and will need to begin by developing a master sample frame for agriculture and an integrated data base. Ethiopia could be grouped somewhere in between group two and three when it comes particularly to agricultural statistics. This is because Ethiopia has a long history of planning and conducting socio-economic and demographic sample surveys on a wider scale. In this connection in the last ten years, the government of Ethiopia has allocated a progressively significant budget for implementing the statistical plans. This was operationally possible due to the establishment of a properly functioning statistical system in the country, under the National Integrated Household Survey Program (NIHSP) by the CSA. Moreover, in order to address the problem of the national statistical system through a more comprehensive approach, the Medium Term National Statistical Program (MTNSP) from 2003/04- 2007/08 was set up and implemented. The new National Strategy for Development of Statistics (NSDS) for the country covers the period 2009/10-2003/14. This statistical strategy differs in content, scope and coverage from the already completed MTNSP. It provides the country with a strategy for strengthening statistical capacity across the entire National Statistical System. This NSDS investigates the strengths and weakness of the statistical system including the agricultural sector. It also identified the priority areas to be considered in the sector and new technologies and methodologies to be implemented in the next five years.

As outlined by the FAO World Program for Census of Agriculture there are several advantages of integrating agricultural statistics into the national statistical system. Major reasons are to avoid

duplicating efforts, prevent the release of conflicting statistics and ensure the best use of resources. In addition, concepts, definitions and classifications become standardized allowing better collection of data across sources. However, Ethiopia has been suffering from divergent crop production forecast estimates that have been provided by various actors until recently. It is expected that once the NSDS is implemented and the CSA could play its leading role, then the advantage of integrated statistical system could be maximized.

3.2 Implement a Master Sample Frame for Agricultural Statistics

The draft global strategic plan put emphasis on the integration of agriculture into the national statistical system which is built on the concept that all census and survey data collections for agriculture be based on sample units selected from a single source, i.e. the Master Sample Frame for agriculture.

As indicated above, the CSA has developed its master sample frame basically from its Population censuses. A master sample frame used for selection of primary sampling units (PSUs) for the integrated household survey is a list of census enumeration areas[3] (EAs). The list contains the size of total, agricultural and non-agricultural households with their geographic information and codes. The second stage sampling units are fresh list of households which are prepared at the beginning of the integrated survey each year.

The strengths and weaknesses of the CSA's master sample frame have been assessed by international consultants. Their findings showed that the master sample frame is a major improvement of the CSA over time to base its NIHSP on probability sampling. The agricultural census as well as annual agricultural sample surveys and all other socio-economic and demographic surveys were based on sample units selected from its master sampling frame. After thoroughly assessing the current system of agricultural statistics of the CSA against their criteria their conclusion was agricultural statistics is: standardized, objective, comprehensive, cost-effective, sustainable, able to detect change and flexible. These criteria were examined separately as follows:

- Standardized – CSA's data are standardized in all regions and zones.
- Accurate - CSA estimates have low sampling errors but possible non-sampling errors; CSA needs better control over the field data collection.
- Timely – CSA estimates must be timelier.
- Objective – CSA data are objective. Objectivity is what gives credibility to a system.

[3] *Census Enumeration Area (EA)* is a unit of land delineated for the purpose of enumerating population and housing units without omission and duplication. An EA usually consists of 150 – 200 households.

- Comprehensive – CSA data are comprehensive; data on crops, social and economic data and nearly any data required by the government.
- Cost-effective – CSA data are cost-effective but this can be improved.
- Sustainable – CSA data system are sustainable.
- Able to detect change – CSA data estimates measure change over time.
- Flexible - CSA data system is flexible meaning new requirements can be added.

On the other hand, the consultants' observation with regards to the limitation of agricultural statistics using the existing master sample frame was the problem associated with under-coverage which might possibly arise due to the following reasons:

- Only rural EAs are sampled; some farmers might live in urban EAs.
- An EA has between 150 and 200 households. It is possible to miss households during household listing.
- Farmers may not "remember" and/or "report" all their fields.
- Measurements of all fields belonging to a selected farmer are may be incomplete.
- Boundaries described in the EA definition could be outdated since population censuses are carried out at least every 10 years and areas might have changed.

3.3 Integrated Household Survey Framework

The vision of the global strategy is that the survey framework starts with the goal that it be sustainable, and provides an annual work program that is consistent from year to year so that the statistical data could fully utilized over time. The complete survey framework includes the sample design, questionnaires, and method of data collection, analysis, and dissemination.

As shown above the agricultural statistics data of the CSA fulfils the vision global strategy. Moreover, the Master Sample Frame that of NIHSP forms the foundation for the integrated survey framework. The framework also considers the need for the linkage between different data items not only between themselves.

In designing the sample strategy for the integrated household survey program of Ethiopia, the following major factors were taken into account:

i. The variability of characteristics.
ii. The multi-subject nature of the survey program.
iii. The average workload of the enumerators.
iv. Simplicity in sample selection, field operations and data processing.
v. The need to integrate various surveys to be able studies the relationships between and among socio-economic and demographic variables.
vi. Available resources.
vii. Level of disaggregation of survey results.

The overall sample strategy was carefully designed for the program to facilitate the integration and linking of different surveys conducted in the same period (year), collect data on each survey as much as possible from the same set of sample of primary sampling units (PSUs) and households (SSUs), adopt the same concepts, definitions and classification systems and incorporate core items in each of the schedules prepared for the conduct of various surveys. The sampling frame used for the selection of PSU is compiled from the national population and housing census.

The existence of one to one correspondence between agricultural holders and households in more than 90 percent of the cases provides an ample opportunity for integrating and linking the agricultural, and other socio-economic and demographic surveys. However, integration and linkage of the different surveys at PSU levels has been granted since data on all topics envisaged in the survey program are collected from the same PSUs. The survey is also designed to link the data at agricultural households (SSU) level.

The geographical and administrative division of the country is used to form the first level of stratification. This has been accomplished through the adoption of the administrative zones as first level stratification. Within each stratum the census enumeration areas (EAs) are used as primary sampling units (PSUs). The secondary sampling units are agricultural households, and with inclusion of some non-agricultural households for other socio-economic surveys.

Therefore, it could be observed that PSUs are selected with probability proportional to size, size being the number of households and overlapping of sampling units had been allowed across surveys to facilitate linkage and estimation of relationships between surveys undertaken in the same or different time periods.

Undertaking integrated household surveys, collecting data from the same set of sample households, with much objectives and topics must have advantages and disadvantages.

The advantages are:

i. The survey forms encompass a wide number and variety of variables and questions which will provide a comprehensive and rich statistical image of household status.
ii. It helps us to undertake more comprehensive and in-depth analysis through statistical intersections between different variables of database which will allow us to highlight the connections and interconnectivity between different variables.
iii. Building aggregated indicators for different socio-economic and demographic areas to be used in the assessment and monitoring of the implementation of the national plans and MDGs.

The disadvantages which can be related to the excessive number of questions in the surveys include the possibilities:

i. Of a high number of refusals to fill the whole form;
ii. The respondents may not complete the whole form due to tiredness (respondents' fatigue);
iii. The respondents may provide wrong or imprecise answers in an attempt to finish the forms as quickly as possible.

Therefore, the integrated household surveys contents are supposed to be determined taking into considerations the possibility of integrating and linking of the different surveys by acquiring the most diversified quantity of data and the respondents' fatigue which may affect the quality of data collected. Administering different modules in different periods (phases) is also a solution to address the problem of respondents' fatigue.

4. Annual Agricultural Sample Survey

The agricultural sample survey is being undertaken on annual basis to obtain basic information on the agricultural sector needed for development planning and socio-economic policy formulation. The sample design is formulated with the aim to obtain estimates of reasonable precision on major agricultural characteristics and farm-gate prices at each reporting domain, regional and national level.

In particular, the objectives of the survey are to estimate:

i. Pre-harvest crop production forecast.
ii. Area under major crops in both main (Meher) and short (Belg) rain seasons.
iii. Post-harvest production and yield of major crops in both seasons.
iv. Land use and practice of modern technologies/inputs by kind and quantity.
v. Area under temporary crops, permanent crops, fallow land, pastoral land and other land use by size of holding.
vi. Number of livestock and poultry by kind, age and sex.
vii. Milk, egg and honey production.
viii. Use of crop and livestock products by type of products.

In order to meet these objectives, the annual agricultural sample survey is actually comprised of several questionnaires/forms administered to households at several times through the main rainy season (*meher*, June-September for planting and from as early as September up to February for

harvest) and the short season (*belg* season from February-May). In the first phase, around September-October, information on each field (sub-parcel) is collected for the 30 sample households including: type of landholding, irrigation, input use, crop conditions. Ten households (of the 30) are selected to also report core crop production forecast information. Cropland area measurement is conducted from October to November in the second phase. In the third phase, households report livestock holdings. In the fourth phase (post harvest), mid October to January, some crop cuts are done and households report agricultural practices and utilization of agricultural and livestock product in the fifth phase (February). The sixth phase (April to June) is during the short rainy season (*belg* season), in the subset of areas with this season; during this visit all households report input use and land area.

As mentioned above, only agricultural statistics data are collected on the annual basis through agricultural sample surveys. Non-farm socio-economic information is not collected in this survey. The Rural Investment Climate Survey (RICS) conducted by the CSA in 2006/07 with support from the World Bank addressed the gap in the information set for rural households. It was conducted among a subset of the agriculture sample survey households and included a booster sample of non-farm rural households. Based on the experience gained from the RICS survey, the CSA planned to conduct the rural economy survey annually in the near future. The rural economy survey is planned to be done in collaboration with the World Bank (WB) Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA) project, which is a new initiative funded by the Bill & Melinda Gates Foundation (BMGF). In addition to the inclusion of the non-farm information, conducting the survey on Computer Assisted Personal Interview (CAPI) and geo-referencing sample households also planned through this collaboration.

The two broad groups that contribute to the nation's agriculture are identified as follows:

1. Private peasant holding
2. Large and medium commercial farms

Accordingly, the data collection techniques applied varies for both groups. In general, however, the collection methods adopted in the conduct of the agricultural surveys covering both groups mentioned have been a combination of interview and objective methods.

With regard to private peasant holdings, an interview is made with the holders within the sampled households, which is the basic unit of enumeration, and thereafter, all fields under sample holder are measured and a sample of fields is taken at random to carry out crop-cutting exercise. In the case of large and medium large scale surveys, information is on yield and

production is obtained through interviews. However, data on acreage size is collected by measuring the area.

5. Quality Assurance of Agricultural Statistics

In most countries the quality of statistical data including data on agriculture is measured using the following eight dimensions. This section describes the quality of the agricultural data provided by the CSA vis-a-vie these dimensions.

i) *Prerequisite of quality*: This dimension deals with the availability of legal and institutional environment, privacy and confidentiality and resource availability. The agricultural data is generated through a sound legal framework set by the government. The CSA is a legally mandated government institution responsible to generate agricultural data. The government is putting adequate funding which enables the CSA to generate agricultural statistics annually. The confidentiality issue was well articulated in the decree 442/2005 and the CSA is handling the confidentiality issue according to the responsibility set to the CSA.

ii) Relevance: This is related to the need to conduct the survey, identification of users need and how well the survey addresses their need. As agriculture is the basis for Ethiopian economy, availing data on agriculture is a top priority. Although it is difficult to say that all the data needs in the agriculture sector are well addressed, the CSA has been trying to provide data based on the priorities. During the usual user-producer forum, an attempt is made to address users' need and their priorities to be covered with the available resource and capacity.

iii) *Accuracy*: This dimension deals with the assessment of sampling errors, coverage, response rates, non sampling errors, data capture errors and source of data. The data is collected from randomly selected agricultural households across the country. The CSA generates agricultural statistics uses what is known as the two stage stratified sampling and it calculates the sampling errors. In addition to the estimates, the CSA usually include the sampling errors in its reports. The non sampling error is addresses through close supervision and data management. A one to four supervisor enumerator ratio is in place to manage the supervision activity in addition to the proper classroom and practical trainings provided each year before launching the survey. A hundred percent verification system is being used during data editing and computer data capturing.

iv) *Timeliness*: This is related to production time, frequency of release and punctuality of release. In the last four-five years, the CSA has shown a significant improvement in the release of its agricultural statistics. In the newly developed NSDS setting a release calendar has also been given a great emphasis.

v) *Accessibility*: The delivery system, metadata availability, presentation of statistics in a meaningful way and means of sharing data are the issues in this dimension. The improvements made in the area of improving data dissimilation system through the improvement of the Information Technology (IT) infrastructure at the CSA address most of these issues. The development of the Ethiopian National Data Archive

(ENADA), the EthioInfo data base, the metadata documentation tool kit and availing data through its website and CD-ROMs significantly improved CSA's data accessibility. Moreover the raw data access policy set by the Council of Ministers has also improved the accessibility of CSA's survey micro-level data.

vi) *Interpretability*: Concepts, definitions and classification and availability of metadata on the methodology used determines the interpretability. The basic concepts, definitions and classification used in the agricultural surveys of the CSA follow international standards mostly generated by the FAO. The concept, definitions and classifications for some of the indigenous variables peculiar to Ethiopia are treated by consulting the Ministry of Agriculture and Rural Development (MoARD). The metadata for the methodology used are usually available with the reports.

vii) *Coherence*: The use of common concepts, definitions, variables and classifications, common methodologies and systems for data collection and processing and common methodology for various processing steps of a survey such as edits within and between series addresses the coherence dimension. The CSA considers this more seriously in order to have a time series data which are comparable. The CSA well realizes the fact that a well set Monitoring and Evaluation system cannot be set without keeping its coherence. As it can been seen from several reports generated, the concepts and the entire procedures of data collection and processing gives a considerable attention to data coherence.

viii) *Methodological soundness*: This dimension deals with international norms and standards on methods, data compilation methods and the soundness of other statistical procedures and techniques employed. The methodological soundness of CSA agricultural data has been witnessed by high level experts in this area several times. Following the development of science and technology, the CSA is implementing new findings in the area to make its methodology more updated.

ix) *Integrity:* The final dimension of data quality is related to professionalism and ethical standards and impartiality. As it is well known, the CSA is an independent entity in generating statistics in the country. All the statistics generated by the CSA including the agricultural statistics are collected, analyzed and disseminated by the CSA's professionals without any interventions.

6. Advanced Methodology Planned to be Implemented at the CSA

CSA is implementing and integrating several advanced technologies. These efforts will improve data the CSA publications by providing data that are more accurate, timely and credible. The sections that follow describe those methodological changes that are being implemented:

a) Area frame methodology
b) Improved stratification
c) Multiple frame methodology
d) Small area estimates methodology
e) Other methodological improvements

6.1 Area Frame Methodology

Definition: An **Area Frame** (AF) is a special case of cluster sampling where farm fields are the clusters. "The concepts of AF sampling are simple: divide the total land area to be surveyed into N small parcels of land, without overlap or omission; select a random sample of n parcels" (Houseman p.1).

The final sampling units (implemented in Ethiopia) are called segments. The process of AF construction is accomplished by selecting the sample in two steps. This two step selection process avoids the necessity of dividing the entire population into small N sampling units.

6.1.1 Survey Design

Area frame design is similar to the current CSA survey design. Enumeration Areas (EAs) are used as primary sampling units (PSUs), which is exactly the same as the current system. With AF sampling, one assigns a segment (about 40 hectare) i.e. secondary sampling units (SSUs) to each EA by dividing its total area of the EA. For example, an EA with 800 hectares will get an assignment of 20 SSUs. This EA now has a size of 20 and that is used in the Probability Proportional to Size (PPS) sampling procedure.

The formal name of the survey design for AF sampling is two stage (or two step) stratified systematic PPS sampling of the first stage of sampling. And an equal probability of selection for the second stage makes the design self-weighting.

6.1.2 Sampling Frame

The sample frame is a list of PSUs that has been changed from EAs with households, to EAs with 40 hectare SSUs. Of course, EAs still contain households but the households do not establish the size, they become a random variable much like wheat and maize. In classical two stage sampling, variance components can be estimated for within and between PSUs. Since only one SSU is selected within each selected PSU, variance components cannot be estimated so statisticians sometimes refer to this design as "two step, one stage sampling."

6.1.3 Survey Sample Size

The sample selection procedure for the AF is similar to what has been used in CSA so far to select EAs. Within each stratum, EAs are listed and the assigned 40 hectare SSUs are accumulated in a column on a spreadsheet. The rows in the spreadsheet are used to identify the EAs. All selected EAs are identified including the random start, the first EA, the random start plus the interval for the second EA and so forth.

Once the PSUs (EAs) are selected, they must be subdivided into the assigned number of SSUs. This is not done in the field but is done in the GIS lab using high resolution satellite imagery and GIS software. A great deal of care is required in this step because the sampling errors can be greatly reduced by careful work at this stage.

CSA has a GIS laboratory with cartographic experts who have been trained to subdivide these EAs into the correct number of SSUs. The SSUs within an EA are numbered and one is selected at random for data collection. The one that is selected will be checked to make sure it has 1) good physical boundaries, 2) approximately 40 hectares, 3) contains some households and fields and 4) belongs to the stratum from which it came because the one selected will represent the stratum from which it came. Because of restrictions on physical boundaries and the resolution of the imagery, a better breakdown is often not possible.

6.1.4 Data Collection

Data collection of the segment is simple in concept. Data must be collected from all reporting units inside the segment boundaries without error. It is not necessary to record farming activities outside the boundaries.

In the AF, the land is selected first and farms are associated with the land. The supervisor and the interviewer have land as a control over the interviewing process. This is in contrast to the current system where farm households are listed and selected and land is associated with the household and no way to control the data collection from the office. The AF procedures can better control field data collection both in the field and office.

Most sample surveys fail in the data collection phase. The AF system in which a segment is identified (either with or without physical boundaries) allows control over the data collection process. Therefore, the last part of this statement appears to be logical, “However, survey

methods differ considerably with regard to effectiveness, or potential effectiveness in coping with practical problems that exist."

6.2 Improved Stratification: Land-cover and Crop Intensity

Stratification in the current EA household list system is at the administrative zones. Every zone has an independent sample from every other zone. While this stratification system is probably useful, zones have tremendous diversity in land use. The new survey design has deeper types of stratification in order for the sampling to be more efficient.

An important part of the construction of an Area Frame described above is the land cover information. In particular the land cover is used for stratification and precedes the area frame construction process. In the case of Ethiopia, Primary Sampling Units (PSUs) or Enumeration Areas (EAs) are assigned to land cover classes according to crop intensity.

6.3 Multiple Frame Methodology

Multiple frame sampling is an advanced type of technology that should always be used with area frame methods. Multiple frame (MF) sampling employs the use of two or more sampling frames conjointly. Collectively, frames should include all farms in Ethiopia. Some farms will be included in two or more frames. The most common application of MF sampling is two-frame sampling. One frame is a general purpose frame such as an AF in which all farms are included. The second frame is a list of farms that are important farms for a specific item of interest. In Ethiopia, the example where an AF is used for all farms and a list frame of medium and large scale commercial farms is a good example of correct usage.

6.4 Small Area Estimates

6.4.1 Current Reporting Level

With probability estimates and controlled non-sampling errors, the higher the level, the more accurate the data. For example, the national level data are the most accurate and are established first because the sample size is largest. The national estimate is divided into regional estimates and these are less accurate than the national estimate because the sample sizes are smaller. Nevertheless, these are still good estimates but when regional estimates are made, these regional estimates must add back to the national estimates. Within regions are zones and zonal estimates must add back to regional estimates but within zones direct expansion estimates are not made because sample sizes are too small.

6.4.2 Reporting at Woreda (District) Level

Estimates are not made by direct expansions for woredas. However; woreda level estimates are important to the Regional State Governments. One way to provide these woreda level estimates is to increase the sample size; however, increasing the sample size in order to provide woreda level estimates is not feasible. Cost, time and accuracy would be affected adversely if the sample size is to be increased to accommodate woredas level estimates annually unless and otherwise it is agricultural census that should be conducted once every ten years. A cost-effective solution is to use auxiliary information from other sources and apply the technique of small area estimation (SAE). Fortunately, at present there are a number of reliable techniques available for SAE and further, there are sources of auxiliary information.

The CSA has been trying to generate small area (wereda level) data for agriculture using small area estimation technique. Three different types of data have been used as an input for the small area model. These are data from the last (2001) agricultural census, wereda level data from the MoARD and the direct estimates from the annual agricultural sample survey of the CSA. Evaluations at different levels have been conducted to examine the validity of the data obtained from small area model results prior to dissemination. This exercise of checking and evaluating the data will also continue to improve the result.

7. Concluding Remarks

In the overall assessment, the CSA has been and is still producing consistent data over time. With efforts being made to improve the reliability and timeliness, one could say agricultural statistics in Ethiopia is relatively in a better situation comparing to many developing countries.

With financial support of FAO/EC Project fund, the CSA and MoARD experts had a chance to do study visits in various countries, namely: Korea, Thailand, Morocco, and Indonesia, to benchmark in improving agricultural statistics and harmonizing the efforts of both institutions in producing relevant data for improving food security information. In general, the CSA learned important lessons from the benchmarking study visit in which the main points could be summarized as follows:

- How to improve crop yield estimates by revisiting the size of crop cutting plots based on physiological and seeding feature of specific crops,

- How to improve crop production forecast estimates using econometric models,
- The Advantages of utilizing multiple frame i.e area frame and list frame to improve area under crop,
- The advantages of introducing portable computers for data collection to improve timeliness of data dissemination,
- More importantly, the benchmarking study to Korean National Statistics Office clearly showed that how the CSA could play its leading role in data quality assessment framework. The data assessment will be applied to all statistics produced by different institutions as well as to that of the CSA.
- Lessons learned from Morocco are also important in shading an insight as to how the remote sensing data and satellite imagery could be utilized in the construction of area frame.

Once the advanced methodological changes and advanced technology mentioned above are completed and the lessons learned are implemented, the problems and the questions raised by many data users could easily get solutions and timeliness issue could be improved dramatically. It is to be noted that the CSA is currently in the process of establishing the Ethiopian Data Quality Assessment Framework with the financial and technical supports of the World Bank which will be materialized in a near future in collaboration with other stakeholders.

References

1. CSA, 1983. The Rural Integrated Household Survey Program: *Methodological Report, for the Period of 1980 - 83*. Statistical Bulletin 34. Addis Ababa.

2. CSA, 2010. Agricultural sample survey, 2009/2010: *Report on Area and Productions of Crops (Private Holdings Meher Season)*. Statistical Bulletin 446, Volume I. Addis Ababa.

3. CSA, 2009. *Agriculture Statistical Methodology in Ethiopia*. Addis Ababa.