

Developing a master sampling frames for integrated agricultural and rural statistics: the case of Ethiopia

Aberash Tariku Abaye;

Central Statistical Agency;

Addis Ababa, Ethiopia;

Email: [kaberash@yahoo.com](mailto:kaberash@yahoo.com)

Abstract

Master sampling frame is a basis for all socio economic surveys including agriculture. Appropriate variables needs to be collected during census to develop good master sample frame for agriculture. As the master frame is basically developed from data collected during census, very serious attention should be given for master frame development for agricultural survey during census. Appropriate frame; that is list, area or multiple needs to be implemented accordingly. It will also be good to collect community frame to further verify and validate the master frame and also to get appropriate stratifying variable. New technologies like scanning may save time for frame compilation, but sustainable capacity on the technology should be developed. Research needs to be done to save time in master frame data compilation and also to improve the data quality.

Key words: Area frame, multiple frame, community frame

## 1. Introduction

The major duties and responsibilities of Central Statistical Agency of Ethiopia (CSA) are to collect, process, analyze and disseminate statistical data, to provide technical guidance and assistance, and to coordinate the statistical system so as to maintain statistical data quality.

The Central statistical agency of Ethiopia has developed the National Strategy for the Development of Statistics (NSDS) to be implemented for five years. The NSDS has six themes. Development of master frame is one of the major activities of theme 4 in the NSDS. Activities which help for better integration were also designed and implemented in the NSDS. Some of these are Coordination unit established to facilitate implementation of the NSDS, Standard concepts and definitions and industrial and labor classifications are prepared and disseminated, the data quality assessment frame work has been developed, Statistical activities are supported by GIS and IT, a preparatory activity to develop common data bank is started.

CSA has been collecting agricultural data through annual agricultural survey for more than three decades in integrated approach. The major types of agricultural data collected by CSA are cultivated area and production by crop type, land utilization, crop utilization and agricultural practices. For collecting agricultural data, list frame approach for the ultimate sampling unit is used. The list of enumeration areas (EAs) from population and housing census is used as primary sampling unit (PSU) and list of households prepared during the survey in the sampled enumeration area used as secondary sampling unit (SSU). Data from sampled households is collected by both objective measurement and interview method. Expert judgment is also used to support the data. Integrating surveys helps to save cost and also to link data for in-depth analysis. But it may create some difficulty since different surveys may need different measure of size and also different sample size. The agricultural data is reported at zone level. Policy makers require data at lower district level. Such data can be obtained by agricultural census. CSA conducted research to implement ratio and model based method to get data at lower level.

CSA is conducting different research activities to improve agricultural data quality. This includes research on area frame approach which is expected to result in more timely and accurate data. In the area frame approach, Enumeration areas are used as PSU and segments of size 40 hectare are used as secondary sampling units. The land cover classification is used as a base for the development of area frame. In this approach commercial farms data is collected by using list frame, hence multiple frame is applied.

Community level data was collected during population and housing census and used to develop community data frame. These data was collected at EA level.

## 2. Developing master frame for integrated agricultural and rural statistics: listing frame approach

Central Statistical Agency of Ethiopia (CSA) has developed a master frame which can be used for sample selection from population and housing census. A listing questionnaire was designed to collect data to be used for master frame development and the data was collected at the beginning of population and housing census. The major variables collected in the questionnaire are agricultural households, non agricultural households, total households, small and large scale manufacturing, cottage industry, whole sale trade, retail trade and service trade.

The listing frame was scanned. Once the scanning is completed, key correction was also done for the scanned data. Finally the scanned data was exported. The exported data was cross checked with the actual census file. The document for the mismatched EA was identified from the documentation and rescanned and merged with the original file. The frame compilation, which is

exporting the scanned data and cross checking with the census id was done after the census result is released. This creates difficulty in matching.

The frame with total number of households was compiled directly from the census questionnaire and CSA is using that frame now. For large scale farms, commercial frame is compiled. Farms classified as commercial are state farm, private commercial farm, cooperative farm and enterprise farm. This frame is updated every year.

The master frame developed from population and housing census is used to select samples for annual agricultural survey. The primary sampling units which are the enumeration areas are selected by probability proportional to size, size being the number of households from population and housing census. Households are then listed at the beginning of the survey and sample households are selected. Once the sample households are selected, then data on cultivated area and production by crop type, land and crop utilization and agricultural practices are collected. The same sample households are used to collect data for long rainy season, short rainy season, forecast and livestock.

CSA uses integrated approach for data collection. The enumeration areas selected for agriculture survey are also used to collect data for other household surveys such as household income survey and welfare monitoring survey. This integration between surveys helps to link the data for in-depth analysis. The master frame developed by CSA is also used by other sector ministries and research institutes to conduct a survey. The procedure here is ; the sector ministry submits its request to the CSA and then CSA selects sample and provides the list of sampled EAs with their maps to the sector ministry. This facilitates standardization and minimizes the data discrepancy.

CSA is planning to conduct agricultural census in 2015. The enumeration areas selected for agricultural census will then be used as a frame for annual agricultural survey to be conducted after the agricultural census. The first agricultural census was conducted in 2001. This agricultural census was large sample census which provides data at lower district level. The master frame developed from the previous population and housing census used as a frame for the first agricultural census. Then the agricultural census was used as a master sample frame for annual agricultural survey until the recent population and housing census was conducted.

The list frame facilitates integration between different household surveys, good to collect socio economic data and helps to have diversified representation. This method does not require imagery. The difficulty with list frame is that it is time consuming as the households are distributed in the enumeration area. Some holdings can be missed as their holding is dispersed. It makes supervision difficult. In list frame indirect measure of size, which is number of households is used for agricultural survey. The difficulty to update the master frame due to frequent administrative boundary changes is also a challenge. A simplified mechanism for master frame updating is very important.

To generalize, it is a good practice to collect data to develop master sample during population and housing census. Attention should be given to the quality of frame as it is a base for different surveys to be conducted until next census. Appropriate editing mechanism should be set to improve the quality of frame. It is advisable to do frame compilation in parallel with the census data compilation so as to cross check together. Serious attention should be given for documentation and the capacity to scan, do key correction, export and merge the data has to be built well and should be sustainable. More simplified mechanism to update the master frame needs to be developed.

### 3. Area frame development

#### 3.1 Methodology

In the area frame approach, the observation units are territorial sub divisions instead of list of households/holders/ holdings as in the list frame. The unit of area frame can be points, transects (straight line of certain length) or pieces of territory often named segment. CSA used segment as ultimate unit of area frame. Enumeration areas are used as PSU and segments of size 40 hectare are used as secondary sampling units.

The first step in the area frame survey is frame development. Two inputs, enumeration area maps and land cover map, are used to develop area frame. The EAs which are the PSUs were delineated for the purpose of Population and Housing Census. The criteria to delineate an EA were to have 150 – 200 households in rural areas. Topo-sheet was used as a base map and the GPS readings for the EA corners (turning points) were plotted on the topo sheet and then the EA map was traced from the topo- sheet. The enumeration areas are geo referenced.

The land cover classification activity is designed to produce a land cover data base which will provide a standardized, multipurpose product useful for environmental and agricultural purposes. Satellite imagery, appropriate soft ware and predefined legend are required for the land cover classification. CSA Ethiopia used spot-5 satellite imagery, ARC GIS and MADCAT soft ware for land cover classification. Appropriate legend was derived from the standard legend prepared by FAO. The land is stratified in to land cover categories in the land cover classification (LCCS) which is used as a basis for the development of area frame. The LCCS is used to stratify the primary sampling units (enumeration areas) based on their percent cultivated land.

To develop a frame in the area frame, CSA digitized EA map obtained from census cartographic work is overlaid on the land cover map.

Four strata's are created for the area frame based on crop intensity

- Stratum I      crop intensity 75 % or more
- Stratum II     crop intensity 50 to 74 %
- Stratum III    crop intensity 25 to 49 %
- Stratum IV    crop intensity less than 25 %

The primary sampling units (EAs) are selected by PPS, size being number of segments. The sampled EA are divided to segments of size 40 hectare and two segments are selected systematically from each EA for data collection.

Closed segment approach is used for data collection, i.e. all the fields (land use) within the selected segment are listed and questionnaire is filled in for each field. Commercial farms are treated separately and an independent survey is conducted for them. Hence, multiple frame approach which is a combination of area frame and list frame is implemented in the pilot survey.

#### 3.2 Pilot survey conducted in area frame approach

##### West shoa pilot survey

- area frame methodology was started in one zone, west shoa zone of oromiya region
- 40 enumeration areas and 40 segments (1 segment per EA) were sampled

- This survey helped just to check the practicality of area frame. It served like pre test.
- Segment map preparation, delineation, listing of fields and filling the questionnaire was tested.

### **The 2010 pilot survey in oromiya region**

- In 2010 E.C pilot survey conducted in all zones of oromiya region
- By incorporating the recommendations in the west shoa pilot, the 2010 pilot covered all zones of oromiya region.
- 215 EAs and 430 segments selected (2 segment per EA)
- The sample allocated to all zones proportionally
- In this pilot improvements in implementing the area frame approach was observed
- Looking at the results of the survey, significant differences in the estimate of area b/n the list frame and area frame occurred.
- Identification of possible sources of the differences was done by international consultants and recommendations made for the 2011 pilot survey.

### **The 2011 pilot survey in oromiya region**

The recommendation in the 2010 pilot survey implemented

- To clearly measure part of field only with in the segment boundary, when the boundary dissects the field
- To prepare the clear manual and give intensive training
- Allocate enough time for data collection
- Prepare clear segment map with clear boundary as much as possible
- To avoid segments of less than 2 % crop land in stratum 4 in nomadic areas
- To identify and exclude commercial farm
- 239 EAs and 478 segments sampled
- Estimate of major crops are found to be comparable for list and area frame except sorghum, coffee and chat
- Stratum 1 and 2 (high crop land area ) working well but stratum 3 and 4 (less crop land area) creates high variability
- CV's compared and area frame CV are high
- Most of the discrepancies in the estimate of the area and CV occurred due to stratum 4 (less than 25 % crop land stratum)
- It is recommended to further stratify stratum 4 and also to review the issue of some crops.

### **The 2012 pilot survey in four regions in the country**

- As recommended in the previous pilot, stratum 4 is sub stratified in to two
- In the 2012 pilot, three additional regions are covered by area frame pilot survey
- The regions covered by this pilot are Tigray, Amhara, Oromiya, SNNP.
- The pilot data collection is now finalized in Tigray, Amhara, oromiya and SNNP region.
- A final summery for area frame pilot survey is expected from this pilot survey.

Area frame survey saves time as the holdings are near to each other. On the other hand identifying the owner of the field may take some time. It also avoids missing fields. The data can also be cross checked with the total area of the segment. In the area frame approach selecting the appropriate approach for area, production and other socio economic survey should also be thought well. Area frame approach uses appropriate measure of size, which is area. Area frame

approach also facilitates Supervision. One of the disadvantage of area frame is it requires satellite imagery and land cover classification which requires large budget.

#### 4. Development of Community frame

The community questionnaire was designed to collect data on agricultural and health related issues such as agro ecology, types of crops grown in meher (long rain season) and belg (short rain season) season, culture of growing rare crops, irrigation and types of major diseases in the enumeration area. The data was collected only from rural areas during population and housing census. The respondents in the community questionnaire were representatives in the enumeration area which are from 3 to 5 peoples composed of elders, officials and development agents

After the data is collected, the identification part of the community frame was edited. This facilitated the frame compilation very highly. For the community questionnaire, the data was captured by data entry instead of scanning. The data entry was done after editing and this minimizes the mismatch. The community frame compilation is a good start.

The community questionnaire is basically useful for stratification purposes for surveys. The information collected in community questionnaire is based on respondent's opinion; and this may create discrepancy with the data collected on scientific basis.

#### 5. Conclusion and recommendation

- Master sampling frame is a basis for all socio economic surveys including agriculture
- As the master frame is basically developed from data collected during census, very serious attention should be given for master frame development for agricultural survey during census. Appropriate variables needs to be collected during census to develop good master sample frame for agriculture and rural statistics.
- Appropriate frame; that is list, area or multiple needs to be implemented accordingly
- It will be good to collect community frame to further verify and validate the master frame and also to get appropriate stratifying variable
- Developing master sample frame from agricultural census should also be considered
- Simplified frame updating mechanism should be studied.

#### Reference

Food and Agriculture Organization of the United Nations (FAO). (1998) "multiple frame agricultural surveys volume 2", FAO statistical development series 10

Aberash T. (2010) "Combining enumeration area maps and satellite images (land cover) for the development of area frame (multiple frames) in an African country: preliminary lessons from the experience of Ethiopia, ICAS V, Uganda

Aberash T. (2011) "Comparative study between area frame and list frame for agricultural survey", 22<sup>nd</sup> AFCAS , Ethiopia