Small Area Estimate in developing countries – the case of Ethiopia

Aberash Tariku Abaye

Central Statistical Agency of Ethiopia

kaberash@yahoo.com

Agriculture is the base of the economy in developing countries. The majority of the population lives in the rural part of the country. The agriculture sector has significant contribution for the country's GDP and also employs large proportion of the population.

Quality statistics is needed for evidence based planning and decision making. The availability of data at lower level like district is essential as the practical planning and decision is made at that level. Conducting agriculture census frequently to get data at small area level is not feasible and economical for developing countries like Ethiopia. Hence the application of small area method to produce data at lower administrative level is an ideal solution for developing countries.

In Ethiopia, Model based district level crop area estimation is done on pilot level. Two auxiliary variables; routine data from ministry of agriculture and the agriculture census, are used in addition to the annual agricultural sample survey data. Area level model (Fay and Herriot model) is used for estimation. Area level model chosen based on availability of data.

Key words: Small area estimation, Agriculture census, direct estimate, administrative data

1 Introduction

Agriculture is the base of the economy in developing countries. The majority of the population lives in the rural part of the country. The agriculture sector has significant contribution for the country's GDP and also employs large proportion of the population. Quality statistics is needed for evidence based planning and decision making. The availability of data at lower level like district is essential as the practical planning and decision is made at that level. Conducting agriculture census frequently to get data at small area level is not feasible and economical for developing countries like Ethiopia. Hence the application of small area method to produce data at lower administrative level is an ideal solution for developing countries.

Most sample surveys conducted by statistical agencies provide accurate statistics at a higher level. But policy makers are interested in obtaining statistics for smaller domains/small areas.Those surveys planned to get estimate at aggregated level do not contain sufficient sample sizes to give reliable estimates by themselves to small areas. Currently different methods of estimation in small areas which incorporate auxiliary information from other sources are available.

Ethiopia is a country structured in to 9 regional states and 2 city administrations. Totally around 96 zones and 1064 districts (woredas) are found in the country. The annual agricultural sample survey which is conducted by Central Statistical Agency (CSA) of Ethiopia is planned to provide reliable estimate at zonal level. But there is an increasing request to produce woreda level estimates. Since the annual agricultural sample survey is planned to produce zonal estimate, the direct estimates for woredas which we get from this survey is not reliable as the sample size for woredas is small. Hence, small area estimation method was tested in collaboration with FAO. Two auxiliary variables, data from ministry of agriculture and data from agricultural census were used in model based estimation for woredas. The data from the ministry of agriculture is the administrative data compiled by the development agents and aggregated at woreda level where as the census data is the data obtained from the agriculture census conducted by CSA which provides data at woreda level.

2. Objective of the small area estimation

The objective of small area estimation in Ethiopia was to get reliable estimate at woreda level based on CSA zonal estimate and two auxiliary variables, ministry of agriculture data and agricultural census data. The purpose of this activity was to get model based woreda level crop area estimation. The small area exercise was used only for internal consumption with in CSA.

3. Methodology used in Small area estimation for crop land Area

Area level model (Fay and Herriot model) is used for model based wereda level crop area estimation. Woredas are considered as small areas. Area level model is chosen based on availability of data. Empirical best linear unbiased predictor (EBLUP) estimator used. The models were fitted at regional level as the number of observations is higher and the CV is better at regional level than the zones.

It may be noted that only some of the woredas (m) are represented in the sample. For the woredas not represented in the sample, it would not be possible to develop direct estimators and hence regression synthetic estimator is used. Two different procedures used for estimating Mean Square Error for sampled and non-sampled woredas.

The estimation procedures were coded in R software and used for estimation. Woreda level estimates are developed only for important crops for which reliable estimates at regional/ zonal level are available. The six major crops which are selected based on the importance of the crop and CV are

Teff, Barley, Maize, Sorghum, Wheat and Finger millet. The EBLUP estimates were done for five major regions. The estimation was not done for regions where the CV is high and majority of the population is nomadic.

Appropriateness of the input variables was checked first. The major four assumptions of linear model were checked for the fitted regression model. Linear relationship for the input variables observed with some indications of the outliers. The R^2 value indicates that additional covariates need to be incorporated in the model.

Evaluation of the results was also conducted. The measure of bias, model misspecification and spatial randomness were checked. The EBLUP estimate was found as unbiased and the pattern of the residuals was spatially random. The mean square error for each estimate was calculated and it was observed that the CV for EBLUP estimate is better than the CV for direct estimate. But the CV for woredas which does not have direct estimate was found to be high. To maintain additivity to larger area, calibration of the EBLUP estimate needed. Hence the EBLUP estimate is adjusted with the CSA zonal estimate.

4. Challenges observed

The woredas administrative boundary changes affect the exercise significantly. Merging and splitting of woredas (Districts) happen frequently. In such cases it was difficult to link the agriculture census data, ministry data and annual survey data. To address this; imputed data using proportions was used.

Automating the procedures was another challenge. The procedures were coded in R software. More simplified automated procedures are required to be used by developing countries.

Estimation of MSE for non-sampled woreds was complicated. It creates difficulty in preparing codes in R. The MSE was also very high for these woredas. Over sampling was done for the following year annual agricultural survey so as to get direct estimate for all woredas. Standard methodology to improve the estimates from non-sampled woredas should be designed.

Collecting and organizing data from the ministry, agriculture census and annual survey was time consuming. Better coordination among stakeholders will facilitate the data collection. The quality of the data especially from administrative records also needs to be improved. Design and implementation of the National Strategies for the Development of Statistics (NSDS) in general and Strategic Plans for Agricultural and Rural Statistics (SPARS) in particular may facilitate the coordination and also to improve the quality of data.

Capacity to design the methodology, run the model and prepare reports is needed. And more simplified mechanism for checking the validity of the estimates should be available.

5. Conclusion

Small area estimation technique is very useful for developing countries like Ethiopia to produce estimates at lower administrative level. It will facilitate informed planning and decision making. More organized methodological exercise to better guide developing countries on the implementation of small area technique is important. This may include provision of simplified automated procedure for easy implementation of the small area exercise and also support countries to improve the coordination among the members of the National Statistical System (NSS) so as to improve quality of data. Above all, capacity building for developing countries in the small area estimation technique is very important.

Reference

1. **Rao, J.N.K., small area estimation**
2. **A guide to small area estimation by Australia bureau of statistics**
3. **Diagnostics for small area estimation, Statistics Newzeland.**
4. **Evaluation of small area estimation methods , An application to unemployment estimates from the UK LFS. Gary Brown, Ray chambers, Patrick heady, Dick Heasman.**
5. **General Linear Models small area estimation: An assessment in agricultural surveys. Carlo Russo, Massimo Sabbatini and Renato Salvatore, university of casino, Italy**
6. **Statistics Canada International Symposium Series – proceedings, symposium 2003: challenges in the surveys taking for the next decade**
7. **Goodness of fit tests for mixed model diagnostics, jiming Jiang**