Testing of the AGRIS methodology in Ghana

Note on Listing and Sampling

1 Sampling design of AGRIS-Ghana pilot survey

1.1 Background

The objective of the pilot survey is to collect reliable data using AGRIS questionnaires customised in the Ghanaian context. This survey should allow reliable estimates of the different AGRIS modules indicators related to major agricultural performances, labour, economic activities, environment machineries and others agricultural statistics.

1.2 Estimation domains

The estimation domain chosen for this activity is the district. Therefore, this survey should allow reliable district-level estimations. It has been decided to collect in each selected district the core module of AGRIS plus one of the four rotating modules. Therefore four districts will be covered by the survey. The Ghana Statistical Service (GSS), the main implementing partner, has already selected its districts of interest in the Ashanti region for the pilot survey: Ejura Sekyedumase, Ahafo Ano North, Asante Akim North and Sekyere Afram Plains North.

1.3 Statistical units and sampling frame

The observational and sampling units of the pilot survey are the agricultural holdings covering both the household and the non-household sectors as adopted by AGRIS. Given that there is no updated list of agricultural households in the country, the complete list of enumeration areas in the chosen districts from the 2010 population housing census (PHC) will be used as a frame for the holdings in the household sector. By definition, the holdings in the non-household sector are economic units such as corporations and government institutions engaged in agricultural production. A complete list of these holdings in the chosen districts should be provided by GSS and the ministry of agriculture (MoFA) to be used as a sampling frame for this category of holdings. The sampling frame is therefore a multiple frame composed by the two lists mentioned above for the two categories of holdings.

1.4 Sampling design

As recommended in the AGRIS sampling strategy, a stratified two-stage sampling design will be used for the holdings in the household sector. The primary sampling units (PSU) are the enumeration areas (EAs) and the secondary sampling units (SSU) are the agricultural households. A stratified simple random sampling will be used for the holdings of the non-household sector.

1.5 Sample size

Holdings in the household sector

For these holdings, the calculation of the sample size is performed through the approach based on the precision required of the final estimates for the main variables of interest. The main variable of interest considered is the area of the agricultural land owned by the households. This information has been collected during the 2012-2013 Ghana Living Standards Survey 6 (GLSS6). Therefore the data of this survey has been used to calculate the coefficient of variation (CV) of the variable of interest in the chosen districts.

It should be noted that the estimation domain of the GLSS6 was the region. A two-stage sampling design has been used in that survey and the primary sampling units which are the enumeration areas have been selected in each region with the probability proportional to size (PPS). The measure of size is the number of populations. Fortunately all the chosen districts for the AGRIS-Ghana pilot survey have been covered by the GLSS6. For the calculation of the CV of the households' agricultural land, it is assumed that the EAs sampled in GLSS6 and located in the target districts have been selected in these districts with the same method of selection (PPS). Thus the households covered are supposed to have been selected in the districts with a two stage sampling design.

The classical formula below is used for the calculation of sample size of households:

$$m_d = \widetilde{D}_{eff} \times \frac{1}{g} \times \frac{CV_{yU_d}^2}{CV^{*2} + CV_{yU_d}^2/N_d}$$

N_d is the total number of households in the district d from the 2010 PHC.

$CV_{yU_d}^2$ is the CV of the households' agricultural land (y) in the district d computed from GLSS6 data.

CV^{*2} is maximum relative error accepted for the survey (10%).

$\widetilde{D}_{eff}$ is an estimate of the design effect. According to Petterson and do Nascimento Silva (2005), in developing countries, the two-stage selection of households provides a design effect that can range from 2 to 6. The maximum value 6 has been used here.

g is the expected response rate calculated from the GLSS6 data for each district.

The number of households to be surveyed in each PSU is fixed to 10. Therefore, the size of the sample of PSU is the size of the sample of the households divided by 10.

District	Holdings in the household sector		
	SSU (Holdings)	PSU (EAs)	Final SSU sample
Ahafo Ano South	55	6	60
Asante Akim North	141	14	140
Ejura Sekye Dumasi	71	7	70
Sekyere Afram Plains	104	10	100
Total	**371**	**37**	**370**

Holdings in the non-household sector

These holdings include:

- Registered commercial farms (corporations, modern farms…)
- Informal commercial farms
- Government farms (e.g. farms of the agricultural research institute, farms of schools, prisons or other organisations)

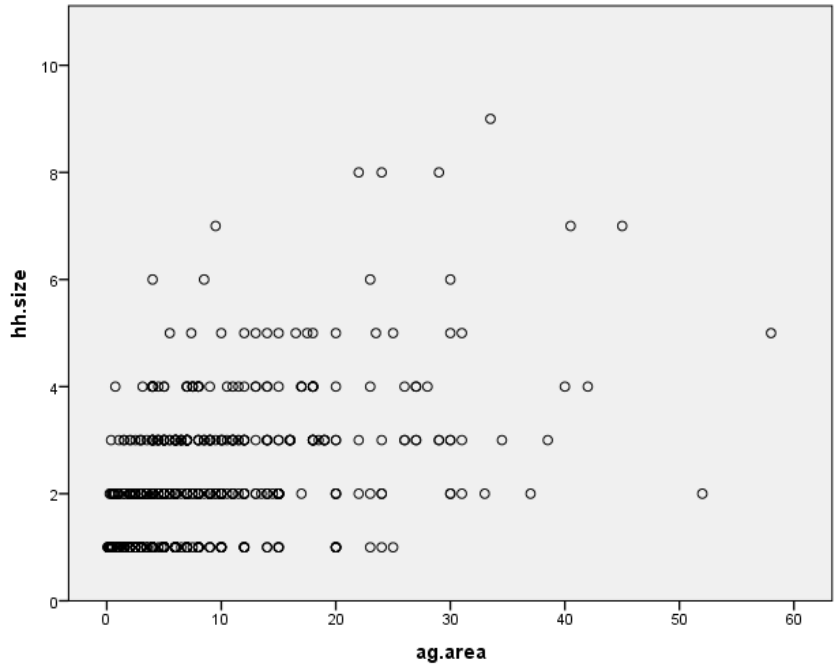
Initially, GSS and MoFA were requested to provide a complete list of holdings in the non-household sector to be used as sampling frame. Therefore, the plan was to use as overall sampling frame a multiple frame composed by the two lists described above (one for the household sector and one for the non-household sector). However, after further discussion, it was realized that the list of holdings in the non-household sector received could not be considered as a reliable sampling frame for the targeted units. As a consequence, even though data for 80 households in this list was collected, it was decided not to analyse data for holdings in the non-household sector.

1.6 Sample selection

The holdings in the non-household sector will be selected in each stratum through a simple random sampling.

For the agricultural households, the PSU will be selected with a probability proportional to size and 10 SSU will be selected through a simple random sampling in each PSU. The PPS sampling can reduce considerably the variance of the estimator of the total if the measure of size is well correlated to the main variable of interest. As stated above, here the measure of size is the number of populations and the main variable of interest considered is the area of the agricultural land. The correlation between the two variables has been calculated using GLSS6 data. As it can be seen in the graph and the table below, there is a quite good and significant correlation between these variables.

Correlation



Correlations				
			ag.area	hh.size
Pearson Correlation	ag.area	Correlation Coefficient	1	0.557**
		Sig. (2-tailed)		0.00
		N	815	815
	hh.size	Correlation Coefficient	0.557**	1
		Sig. (2-tailed)	0.00	
		N	815	815
Spearman's rho	ag.area	Correlation Coefficient	1	0.590**
		Sig. (2-tailed)		0.00
		N	815	815
	hh.size	Correlation Coefficient	0.590**	1
		Sig. (2-tailed)	0.00	
		N	815	815
**. Correlation is significant at the 0.01 level (2-tailed).				

1.7 Stratification

Holdings in the household sector

In a two-stage sampling, a great proportion of the total variance comes from the inter PSU variance. Therefore in order to improve the estimations, a size based stratification of the PSU has been performed. The generalized Lavallee-Hidiroglou (1998) method has been used for an optimal stratification.

District	Nb EAs	Nb strata
Ahafo Ano South	215	5

Asante Akim North	114	3
Ejura Sekye Dumase	162	4
Sekyere Afram Plains North	64	2

Sample of EAs for the Holdings in the household sector

District	Stratum code	EA Code	EA Name
Ahafo Ano South	1.1	616100008	Manhyia
Ahafo Ano South	1.1	616100112	Adugyama
Ahafo Ano South	1.2	616100156	Kwamekyemkrom
Ahafo Ano South	1.3	616100087	Domeabra
Ahafo Ano South	1.4	616100007	Akyease
Ahafo Ano South	1.5	616100141	Asuoadei
Asante Akim North	2.1	629100112	Domeabra
Asante Akim North	2.1	629100133	Hwidiem
Asante Akim North	2.1	629100164	Agogo
Asante Akim North	2.1	629100182	Agogo
Asante Akim North	2.1	629100206	Kasanso
Asante Akim North	2.2	629100116	Domeabra
Asante Akim North	2.2	629100140	Amentena
Asante Akim North	2.2	629100156	Agogo
Asante Akim North	2.2	629100183	Agogo
Asante Akim North	2.2	629100200	Agogo
Asante Akim North	2.3	629100132	Hwidiem
Asante Akim North	2.3	629100151	Wioso
Asante Akim North	2.3	629100180	Agogo
Asante Akim North	2.3	629100198	Agogo
Ejura Sekye Dumase	3.1	626200033	Nokwareasa
Ejura Sekye Dumase	3.1	626200155	Abrewa Ano
Ejura Sekye Dumase	3.2	626200128	Ebuom
Ejura Sekye Dumase	3.3	626200066	Ejura
Ejura Sekye Dumase	3.3	626200112	Sekyedumase
Ejura Sekye Dumase	3.4	626200050	Ejura
Ejura Sekye Dumase	3.4	626200093	Ejura
Sekyere Afram Plains North	4.1	630100129	Magyefa
Sekyere Afram Plains North	4.1	630100159	Mfutrom - Konkonba
Sekyere Afram Plains North	4.1	630100172	Saabum
Sekyere Afram Plains North	4.2	630100122	Wiribongtire
Sekyere Afram Plains North	4.2	630100127	Mamprusi No.I
Sekyere Afram Plains North	4.2	630100136	Funsua (Afoasua)
Sekyere Afram Plains North	4.2	630100147	Abenoa
Sekyere Afram Plains North	4.2	630100151	Kwabena Adu
Sekyere Afram Plains North	4.2	630100157	Drobonso
Sekyere Afram Plains North	4.2	630100176	Anyinofi

1.8 Estimation procedures and sampling errors

A stratified two-stage sampling design was used for holdings in the household sector with the enumeration areas (EAs) as PSUs and agricultural households as SSUs. Below we use and adapt notations and formulas from the AGRIS Handbook (GSARS, 2018).

Notation

h = stratum

H = total number of strata

i = PSU

N = total number of PSUs

I_h= total number of PSUs in the *h*-th stratum

j = SSU

M_{hi}= total number of SSUs found in the *i*-th PSU in stratum *h* ($j = 1, 2, \dots, M_{hi}$)

$M = \sum_h \sum_i M_{hi}$= total number of SSUs in the country

F_{hi}= total number of SSUs listed in the sampling frame as belonging to the *i*-th PSU in stratum *h*

$F_h = \sum_i F_{hi}$, is the total number of SSUs listed in the sampling frame in stratum *h*

n_h= number of sample PSUs selected in stratum *h*

m_{hi}= number of sample SSUs selected in *i*-th PSU in stratum *h*

y_{hij}= value of the target variable *Y* observed on the *j*-th SSU, in *i*-th PSU in stratum *h*

Estimators

The probability of selecting the SSU *j* in the sample is the product of the probability of selection of the PSU *i* in which it is located ($n_h \frac{F_{hi}}{F_h}$) and its probability of selection in the PSU *i* ($\frac{m_{hi}}{M_{hi}}$).

Thus, the *weight* assigned to the SSU f selected in the *i*-th PSU in stratum *h* is:

$$w_{hij} = \left(n_h \frac{F_{hi}}{F_h}\right) * (\frac{m_{hi}}{M_{hi}})$$

Here $m_{hi} = 10$, therefore

$$w_{hij} = \left(n_h \frac{F_{hi}}{F_h}\right) * (\frac{10}{M_{hi}})$$

An estimate of the total amount of *Y* for the entire population may be computed with the following formula:

$$\hat{Y} = \sum_h \sum_i \sum_j w_{hij} y_{hij}$$

The mean of Y can be estimated with two different estimators:

- *Simple mean*

$$\hat{\bar{Y}} = \hat{Y}/M$$

- *Weighted sample mean*

$$\tilde{\bar{Y}} = \frac{\hat{Y}}{\sum_h \sum_i \sum_j w_{hij}}$$

Variance

A simple approximate estimation of the variance of the total is (GSARS, 2018).

$$\tilde{V}(\hat{Y}) = \sum_{h=1}^{H} {M_h}^2 \frac{1}{m_h(m_h-1)} \sum_{i=1}^{I_h} (\hat{Y}_{hi} - \frac{1}{m_h} \sum_{i=1}^{I_h} \hat{Y}_{hi})$$

where $\hat{Y}_{hi}$ and $\hat{Y}_h$ are the estimates of the total amount of *Y* at PSU and stratum levels, respectively.

An approximate estimator of the variance of the mean is:

$$\tilde{V}\left(\hat{\bar{Y}}\right) = \frac{1}{M^2} \tilde{V}(\hat{Y})$$

Coefficient of variation

Total

$$\widetilde{CV}(\hat{Y}) = \frac{\sqrt{\tilde{V}(\hat{Y})}}{\hat{Y}}$$

Mean

$$\widetilde{CV}\left(\hat{\bar{Y}}\right) = \frac{\sqrt{\tilde{V}\left(\hat{\bar{Y}}\right)}}{\hat{\bar{Y}}}$$

2 Note on listings

2.1 Listing 1: holdings of the non-household sector

A complete listing of all non-households farms in each of the four districts was requested without success to MoFA.

2.2 Listing 2: holdings of household sector

After the selection of the samples of the primary sampling units (enumeration areas) in the four districts of interest, the next phase consisted in building a complete list of agricultural holdings in the selected enumeration areas.

The listing was performed by experienced enumerators with an allocation of one enumerator per EA as agreed. GSS will be in charge of the actual field operations.

In each sampled enumeration area, all households (agricultural or not) was listed. The aim was to improve the coverage of all agricultural holdings in the end and allow the comparability of the total population of the EAs with secondary data and data from previous complete listings for quality assessment of the listings.

The listing was performed with a CAPI software: Survey Solutions. A short questionnaire was used to collect basic variables (see paper version of the questionnaire and manual in annex)

Annex
Questionnaire of listing

District code |_| EA code |_|_|_|_|_|_|_|_| EA name: _____

House/Compound No.	Localisation of the Compound	Household Serial No.	Household head				Agricultural practices during the last 12 months							Household old size (number of members)	holder *(person economically responsible for the agricultural production)*			
			name	Age (years)	Sex	Primary occupation	Cultivation of parcels of temporary crops?	Cultivation of parcels of permanent crops?	Number of parcels of crops (all)	Keeping Livestock?	Nb. of Bovine animals	Nb. of Sheep	Nb. of Goats		name	Age (years)	Sex	Highest completed level education
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
					male = 1 female = 2	Farmer = 1 farm hand = 2 fishing = 3 agro-processing = 4 trader = 5 artisan = 6 student = 7 none = 8 other =9	yes = 1 no = 2	yes = 1 no = 2		yes = 1 no = 2 (if no, go to question 15)							male = 1 female = 2	None Pre-school Primary JSS/JHS, Middle SSS/SHS, Secondary Voc/Tech/Comm Diploma Bachelor Degree Higher

Note for enumerators

Respondent
The respondent should be the head of the household. If he/she is away other member of the household and capable of giving all the necessary information may respond.

District code: use the following list (will be prefilled in CAPI)

1 = Ahafo Ano South
2 = Asante Akim North
3 = Ejura Sekye Dumasi
4 = Sekyere Afram Plains

EA codes and names: use the list of sampled EA (will be prefilled in CAPI)

House/Compound code:
Unique number of the house/compound from the sequential numeration of houses/compounds in the EA

Household Serial No
Serial number of the household in the House/Compound

Head of Household: Person acknowledged as such by members of the household and who is usually responsible for the upkeep and maintenance of the household.

- ***Name:*** write the full name
- ***Age:*** Age is to be recorded in completed years
- ***Sex:*** It is important to ask for the sex of the person when information is being given to you by a third person.
- ***Primary occupation:*** occupation on which the head of household spend most of his time

Agricultural practices during the last 12 years: this part collects some information on agricultural practices performed by the household during the last 12 years:

- ***Cultivation of parcels of temporary crops***: ask if the household has cultivated any temporary crops (cereals, vegetable, root/tubers, Leguminous crops, etc.) during the last 12 years
- ***Cultivation of parcels of permanent crops***: ask if the household has cultivated any permanent crops (cocoa, coffee, cashew, banana, Orange, etc.) during the last 12 years
- ***Number of parcels of crops: (temporary and permanent)***: total number of agricultural parcels cultivated.) during the last 12 years
- ***Keeping Livestock? :*** ask if the household kept livestock (Equidae, Bovine animals, Sheep, Goats, Pigs, Poultry, Rabbits, Other animals) during the last 12 years
- ***Number of Bovine animals***: record the number of bovine animals
- ***Number of Sheep***: record the number of sheep
- ***Number of Goats***: record the number of goats

Household size: Record the number of household members (not away from the household for more than 6 months)

Holder: person economically responsible for the agricultural production

- ***Name:*** write the full name
- ***Age:*** Age is to be recorded in completed years

- ***Sex:*** It is important to ask for the sex of the person when information is being given to you by a third person.
- ***Highest completed level education:*** highest education fully completed by the holder