Fafo AIS

Verification of Certification Activities in West African Cocoa Sector

Review of Statistical Weighting Report: Côte d'Ivoire

September 2009

This report describes the findings of a review on statistical weighting report made on scaled-up certification study by the government of Côte d'Ivoire. The review from which the findings are derived was conducted by Fafo commissioned by the International Cocoa Verification Board (ICVB.) The findings of the study offer recommendations on technical aspects of statistical weighting that will refine the quality of the study made on child labor in cocoa production sector.

Anne Hatløy

Tewodros Aragie Kebede

1. Background and materials

In 2007/8 the government of Côte d'Ivoire conducted a certification study with the objectives of estimating child labor and worst forms of child labor (WFCL) in cocoa production sector, as well as documenting the incidence of forced adult labor (PSSTE 2008). Commissioned by the International Cocoa Verification Board (ICVB), Fafo and Khulisa conducted verification activities on the results reported in the study to evaluate the credibility and the scientific basis of the study. Fafo and Khulisa recommended to the ICVB the conditional acceptance of the study. The condition put forward to make the findings of the study credible was that improvements need to be made regarding the estimation of the number and percentage of working children in cocoa sector reported in the scaled-up study. The improvements were needed on these estimates as the inherent design of the data collection did not allow for valid estimations of this key issue.

In addition to the condition that is needed for accepting the report, the verification report also noted that the results reported in the scaled-up study are valid to the selected sample and hence could not be generalized to all cocoa growing households in Côte d'Ivoire. To this end, the verifiers made a recommendation that these results can be made statistically representative and valid for all cocoa growing households in Côte d'Ivoire if the appropriate statistical weights are computed and employed to the results reported in the study. This aspect of the recommendation was implemented by the government of Côte d'Ivoire with a clear objective of enhancing the validity of the scaled-up study and presented a revised report in August 2009. These revised results are the subject of this report in which the findings of the review are discussed. The issue of estimation of working children in the cocoa sector is not addressed in this review.

The materials that are used in reviewing the weighting procedure are:

- PSSTE (2008): National Initial Diagnostic Survey (June 2008)
- CLMS/IDS (2009): Supplemental Study on the Initial Diagnostic Survey

This assessment report is organized by presenting the purpose of statistical weighting in section two. Section three discusses the findings of the review by referring to the technical details of statistical weighting and referring to the appropriate mathematical formulas that should be used for computing weights. Since statistical weighting is a result of the design of the sample survey, close reference is made to the description of the sample design that involves four stages. Each of the four stages is reviewed and the findings are discussed. The final section makes recommendation for the way forward.

2. What is the objective of statistical weighting?

Information on characteristics of populations is constantly needed by various decision makers for planning and designing specific policies that address a given problem. For reasons relating to timeliness and cost, such information is often obtained by use of sample surveys. The scaled-up study in both Côte d'Ivoire and Ghana was conducted to obtain information about children's engagement in cocoa production as well as documenting the incidences of worst forms of child labor and forced adult labor practices. Formally, a sample survey may be defined as a study involving a subset (or sample) of individuals selected from a larger population. Variables of interest are observed or measured on each of the sampled individuals. These measurements are then aggregated over all individuals in the sample to obtain summary statistics (e.g. means, proportions, totals) for the sample. It is from these summary statistics that extrapolations can be made concerning the entire population.

Sampling weights are needed to correct for features in the sample that might lead to bias and other departures between the sample and the reference population. Such features include the selection of units with unequal probabilities, non-coverage of the population, and non-response. Sampling weights are not needed if the characteristics of the sample do not require the application of weights to generalize the results to the reference population. Such cases include when the sample is a simple random sample or the sample is a self-weighting sample. In case of the certification studies, sample weights are needed to generalize the results to all cocoa producing households in Côte d'Ivoire. The objective of weighting that is relevant for the purpose of the certification study in Côte d'Ivoire is the need to compensate for unequal probabilities of selection while other reasons such as non-response can also be taken into account.

The development of sampling weights usually starts with the construction of the *base weight* for each sampled unit, to correct for their unequal probabilities of selection. In general, the base weight of a sampled unit is the reciprocal of its probability of selection into the sample. For example, a sampled unit selected with probability 1/50 represents 50 units in the population from which the sample was drawn. Thus sample weights act as inflation factors to represent the number of units in the survey population that are accounted for by the sample unit to which the weight is assigned. The sum of the sample weights provides an unbiased estimate of the total number of individuals in the target population.

3. Review of weighting procedure: findings

The sample used in the scaled-up study was constructed as a four stage stratified sample, limited geographically to the cocoa producing regions in Côte d'Ivoire. The first stage of the sampling entailed constructing a frame of cocoa producing departments and stratifying them by cocoa production level in 1999/2001 into three categories (low, medium and high) and selecting *departments* based on their production level. The second stage entailed random selection of two *sub-departments* within the selected departments. The third stage entailed random selection of one *village* within each sub-department according to the production level of the category. The last stage involved selection of a fixed number of households from each village. Development of sampling weights usually starts with the construction of the *base weight* for each sampled unit, to correct for their unequal probabilities of selection. The base weights must reflect the probabilities of selection at each stage. In reviewing the computations of weights outlined in the supplemental study (CLMS/IDS, 2009), we followed the design stages and identified the right formulas that should be used for calculating the probability of selection of the unit in a given stage. The assessment of the four stages is presented as follows.

First Stage: No Error

The 51 cocoa producing *departments* in Côte d'Ivoire were categorized in accordance with the categorization of cocoa-producing regions, namely:

1. Low cocoa production (2% of national production – 20 departments),
2. Medium cocoa production (11% of national production – 14 departments), and
3. High cocoa production (87% of national production – 17 departments).

In the low-level category, two departments were randomly selected; in the medium-level category, two departments were randomly selected; in the high-level category, all the 14 departments were selected, except three departments that had been visited during pilot. Altogether 18 *departments* were sampled. The inclusion probability of *departments* is given by:

$$p_{c,\,d} = \frac{n_{c,d}}{N_{c,d}}$$

where $n_{c,d}$ is the number of departments sampled in each category of cocoa producing area (c is category and d is department) and is the $N_{c,d}$ total number of departments in each category of

cocoa producing area. This formula was **used correctly** at this stage in the CLMS/IDS (2009) and hence no error was found.

Second Stage: No Error

The second stage involves sampling of the *sub-departments* where from each selected departments, two *sub-departments* were randomly selected. The inclusion probability of sub-departments is given by:

$$p_{c,d,s} = \frac{n_{c,d,s}}{N_{c,d,s}} = \frac{2}{N_{c,d,s}}$$

where $N_{c,d,s}$ is the total number of *sub-departments* in each selected department and *s* is index of sub-department. This formula was **used correctly** at this stage in the CLMS/IDS (2009) and hence no error was found.

Third Stage: No Error

The third stage entails sampling of the one *villages* n each selected sub-department. The inclusion probability of a *village* is given by:

$$p_{c,d,s,v} = \frac{n_{c,d,s,v}}{N_{c,d,s,v}} = \frac{1}{N_{c,d,s,v}}$$

where $N_{c,d,s,v}$ is the total number of villages in each selected sub-department and *v* is index of a village. This formula was **used correctly** at this stage in the CLMS/IDS (2009) and hence no error was found.

Forth Stage: Error found

The final stage involves sampling of 20 households from each selected village. Hence the inclusion probability of a given household is given by:

Equation 1

$$p_{c,d,s,v,m} = \frac{n_{c,d,s,v,m}}{n_{c,d,s,v,q}}$$

where $n_{c,d,s,v,m}$ is the number of households interviewed in each selected village; $n_{c,d,s,v,q}$ is the total number of cocoa producing households in each selected village and m is an index for a household.

The inclusion probability outlined in the CLMS/IDS (2009) is given by:

$$p_{c,d,s,v,m} = \frac{n_{c,d,s,v,m}}{N_{c,d,s,v,m}} \cdot \frac{n_{c,d,s,v,q}}{N_{c,d,s,v,m}} = \frac{20}{N_{c,d,s,v,m}} \cdot \frac{n_{c,d,s,v,q}}{N_{c,d,s,v,m}}$$

However, this formula is incorrect for calculating inclusion probability of a household when it is compared with the appropriate formula describe in Equation 1 above. Hence due to this, an error is made at this stage resulting in an incorrect overall probability of inclusion for a household.

Suggestions for correcting error

The error identified in the last stage can be corrected by applying the right formula. The use of appropriate formula will result in the overall inclusion probability for households that can be given by:

$$p_i = p_{r,d} \cdot p_{r,d,s} \cdot p_{r,d,s,v} \cdot p_{r,d,s,v,m}$$

.

Based on this, the appropriate sampling weight for households and adult workers/children that **should be used** is given by:

$$W_i = \frac{1}{p_i}$$

To provide guidance in correcting the errors, the appropriate formulas are summarized in Table 3.1 together with an indication in what stage the error is made.

Table 3.1 Summary of appropriate formulas and errors found

Stage	Selected unit	Probability of Selection	Notation	Error found
I	Department	$p_{c,\,d} = \frac{n_{c,d}}{N_{c,d}}$	$n_{c,d}$: the number of departments sampled in each category of cocoa producing area $N_{c,d}$: total number of departments in each category of cocoa producing area.	**No**
II	Sub-department	$p_{c,d,s} = \frac{n_{c,d,s}}{N_{c,d,s}} = \frac{2}{N_{c,d,s}}$	$N_{c,d,s}$: total number of sub-departments in each selected department	**No**
III	Village	$p_{c,d,s,v} = \frac{n_{c,d,s,v}}{N_{c,d,s,v}} = \frac{1}{N_{c,d,s,v}}$	$N_{c,d,s,v}$: the total number of villages in each selected sub-department	**No**
IV	Household	$p_{c,d,s,v,m} = \frac{n_{c,d,s,v,m}}{N_{c,d,s,v,m}}$	$n_{c,d,s,v,m}$: the number of households interviewed in each selected village $N_{c,d,s,v,m}$: total number of cocoa producing households in each selected village	**Yes**
Overall probability of selection		$p_m = p_{r,d} \times p_{r,d,s} \times p_{r,d,s,v} \times p_{r,d,s,v,m}$		
Formula for sampling weight:		$W_m = \frac{1}{p_m}$		

4. Summary and recommendations

The objective of this review has been to assess how statistical weighting was employed on the new report released by the government of Côte d'Ivoire (CLMS/IDS 2009). Based on the materials presented, we have reviewed the procedures described in computing statistical weights. Our findings indicate that, an error was made in calculating the statistical weights that resulted from using an incorrect formula in the calculation of probabilities of inclusion for a household at the fourth stage of the sampling procedure. The error introduced at this stage will have an implication in that the overall probabilities and hence the resulting statistical weights will be wrong. It is important to note that in computing the weights, the implication on actual results can be quantified to assess whether it will decrease or increase a given estimate such as the percentage of children working in hazardous cocoa activities. However, such assessment need not be necessary as it is simply an error that is quantifiable and focus should be made to correct the results by using the right formulas. These errors need to be corrected based on the right formulas summarized in Table 3.1. This report did not address additional assessment on the results reported in (CLMS/IDS, 2009) as the identified errors will result in incorrect results.

Based on our assessment on the supplemental report (CLMS/IDS, 2009), the reported results are different in terms of content from the original scaled up certification study report (PSSTE 2008). For the purpose of providing weighted results on the scaled up study, it is sufficient to apply the appropriate weights without having to produce a whole set of different results and additional exploratory analyses.

Finally, it is important to note that verification can be conducted at various stages addressing various aspects of the subject in which the verification is made upon. On technical issues such as statistical weighting, close consultations between verifiers and the governments would have been efficient in terms of allocation of resources needed for conducting the task and the impact this important undertaking has on policy implications. The improvements suggested in this report need to be further discussed with the government of Côte d'Ivoire technical experts to ensure that an understanding is made about conceptual and technical approaches on the issues of statistical weighting as well as the implementation of the weights.