FOOD SECURITY RESEARCH PROJECT

REVIEW OF SAMPLE DESIGN FOR ZAMBIA POST-HARVEST SURVEY (1997/98) AND RECOMMENDATIONS FOR IMPROVING THE SAMPLING STRATEGY AND ESTIMATION PROCEDURES

BY

David J. Megill

WORKING PAPER No. 2
FOOD SECURITY RESEARCH PROJECT
LUSAKA, ZAMBIA
JULY 2000
(**Downloadable at** http://www.aec.msu.edu/agecon/fs2/zambia/index.htm

REVIEW OF SAMPLE DESIGN FOR ZAMBIA POST-HARVEST SURVEY (1997/98) AND RECOMMENDATIONS FOR IMPROVING THE SAMPLING STRATEGY AND ESTIMATION PROCEDURES

By

David Megill
Sampling Consultant

for the Food Security Research Project
Lusaka, Zambia

July 13, 2000

ACKNOWLEDGMENTS

The Food Security Research Project is a collaboration between the Agricultural Consultative Forum (ACF), the Ministry of Agriculture, Food and Fisheries (MAFF), and Michigan State University's Department of Agricultural Economics (MSU).

We wish to acknowledge the financial and substantive support of the United States Agency for International Development (USAID) in Lusaka. Research support from the Global Bureau, Office Agriculture and Food Security, and the Africa Bureau, Office of Sustainable Development at USAID/Washington also made it possible for MSU researchers to contribute to this work.

This study has been made possible thanks to the contributions of a number of people and organizations. In particular, the the Agriculture and Environment Division, Central Statistical Office (CSO), and the Database Management Unit, MAFF, who provided insights in currently applied methodologies regarding Post Harvest Survey enumeration, sampling and estimation. MAFF and CSO also actively participated in technical discussions that took place in Lusaka, providing the opportunity for the author, D.J. Megill (sampling consultant for FSRP), to refine the proposed methodologies and present the draft version of this paper. Substantive contributions to this study have also been provided by members of the FSRP team[1], in particular Cynthia Donovan.

Comments and questions should be directed to the In-Country Coordinator, Food Security Research Project, 86 Provident Street, Rhodes Park, Lusaka; tel: 234 539; fax: 234 559; e-mail: fsrp1@msu.edu

Jan-Joost Nijhoff
In-Country Coordinator
Food Security Research Project/Zambia

[1] The Zambia FSRP field research team is comprised of Jones Govereh, Billy Mwiinga, Jan Nijhoff and Ballard Zulu. MSU-based researchers in the Food Security Research Project are Margaret Beaver, Cynthia Donovan, Thom Jayne, Eric Knepper, Gelson Tembo, David Tschirly, and Michael Weber.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ANNEXES

REVIEW OF SAMPLE DESIGN FOR ZAMBIA POST-HARVEST SURVEY (1997/98) AND RECOMMENDATIONS FOR IMPROVING THE SAMPLING STRATEGY AND ESTIMATION PROCEDURES

By
David J. Megill

1. Background

At the request of the Food Security Research Project and the Zambia Central Statistical Office (CSO), the consultant reviewed the sample design for the 1997/98 Zambia Post-Harvest Survey (PHS), described in the survey reports. Following a review of this sample design, summary results were tabulated from the survey data to examine the distribution of the households for the different domains of analysis such as farm size groups. Summary data were also examined for the different crops and animals.

In order to determine the reliability of the survey estimates based on the current sample design, the CENVAR (Census Variance Calculation System) software was used to tabulate the standard errors for different types of survey estimates and domains of analysis, using the survey data. These results indicate that the coefficients of variation (relative standard error) are high for certain crops which are rare or have a limited geographical distribution.

The recommendations in this report are based on the findings from this review of the sample design and tabulation of standard errors, as well as discussions with Dr. Cynthia Donovan, Agricultural Economist, Department of Agricultural Economics, Michigan State University (MSU). She provided valuable insight into some of the most important issues regarding the PHS sample design and survey results, and coordinated the discussions and feed-back with the CSO staff. Margaret Beaver of MSU provided valuable computer processing support in generating the CENVAR data input file. Jan Nijhoff was also very helpful in obtaining agricultural data files from the 1990 Zambia Census. During the consultant's visit in Lusaka, he had valuable discussions on the PHS sample design and estimation procedures at the CSO with William Mayaka, Assistant Director for Agriculture and Environment, Peter Mukuka, Senior Statistician, Head of the Agriculture Division, and Felix Muchingile, Senior Statistician, as well as the following staff members in the Agriculture Division: Colby Nyasulu, Senior Statistical Officer; Daniel Daka, Statistician; Masikiso Sooka, Statistician; and Joseph V. Chanda, Programmer/Analyst. These staff members provided important feed-back on the draft recommendations for improving the PHS sample design.

The purpose of this report is to document the findings and recommendations from this review. Specific recommendations are made on potential improvements to the sample design. This report also includes the results of the CENVAR analyses, and specifies the procedures for weighting the data and calculating the standard errors. A separate report on "Review of Questionnaire for Post-Harvest Survey (1997/98) discusses issues and recommendations related to the survey questionnaire.

2. Review of Sample Design for 1997/98 PHS

A stratified multi-stage sample design was used for the Zambia PHS. The sampling frame was based on the data and cartography from the 1990 Census of Population, Housing and Agriculture. The primary sampling units (PSUs) were defined as the Census Supervisory Areas (CSAs) delineated for the census. The CSAs were stratified by district within province and ordered geographically within district. A total sample of 405 CSAs was allocated to each province and district proportionally to its size (in terms of households). A master sample of CSAs was selected systematically with probability proportional to size (PPS) within each district at the first sampling stage; the measure of size for each PSU was based on the number of households listed in the 1990 Census. The secondary sampling unit is the Standard Enumeration Area (SEA), defined as the segment covered by one enumerator during the census. One SEA was selected within each sample CSA with PPS for the survey.

A new listing of households was conducted within each sample SEA, and the farm size was obtained for each farm household. The listed households within each sample SEA were then divided into two groups based on farm size: Category A for households with less than 5 hectares (has.) and Category B for households with 5 or more has. The original sampling plan was to select 10 households from each category within the sample SEA, for a total sample of 20 households per SEA. However, it was found that most sample SEAs had less than 10 households in Category B, in which case all of these households were included in the sample with certainty at the final stage of selection. In order to ensure a sample of 20 households within each sample SEA, the remaining households were selected from Category A. Table 1 shows the distribution of sample SEAs and sample households selected for the 1996/97 PHS by province, as well as the number of sample units actually covered by the survey.

Following the data collection for the 1997/98 PHS, it was found that more than 60 percent of the households selected in category B actually had less than 5 has. according to the survey data. This is due to changes in the plans of individual households in the amount of land planted in crops, as well as nonsampling error in the listing data, given the difficulty in obtaining good information on farm size without asking the detailed questions included in the survey questionnaire. Even with the current level of misclassification, this farm size stratification increases the sampling efficiency for producing estimates of total crop area and production. However, it is recommended to try to improve the quality of the listing data on farm size as much as possible. Recommendations on the farm size stratification are presented in Section 4.3 in this report.

Table 1. Number of Sample SEAs and Households for the 1996/97 Post-Harvest Survey, by Province

Geographic Area	Number of Sample SEAs Selected	Number of Sample SEAs Covered	Number of Sample Households Selected	Number of Sample Households Contacted
Zambia	405	383	8,100	7,637
Central	40	40	800	800
Copperbelt	24	24	480	461
Eastern	72	69	1,440	1,380
Luapula	49	49	980	980
Lusaka	14	14	280	280
Northern	80	72	1,600	1,440
North Western	30	26	600	518
Southern	50	50	1,000	1,000
Western	46	39	920	778

The current sample design for the 1997/98 PHS is quite reasonable for a multi-purpose household survey. The selection of the sample CSAs systematically with PPS within each district, and the selection of one SEA with PPS within each sample CSA is an effective sampling scheme. It provides the same level of dispersion as if the SEAs had been selected systematically with PPS at the first stage. At the same time, the sample CSAs provide larger PSUs in which the sample SEAs can be rotated over time. For surveys which measure rare characteristics, it is also possible to list the households in the entire sample CSA when necessary.

Given that a large majority of the rural households in Zambia are involved in agriculture, the sample of farm households is effective for most types of agricultural characteristics being measured by the survey. However, the survey estimates for some crops which are rare or limited to particular geographic areas have relatively high sampling errors. The definition of in-scope farm households for the survey should also examined. Therefore the report recommends certain modifications to the sample design for improving the sampling efficiency for future surveys.

3. Tabulation of Sampling Errors for Estimates from 1997/98 PHS Data

In order to evaluate the effectiveness of the PHS sample design in meeting the survey objectives, it is first necessary to measure the level of precision for the survey estimates based on this design. Therefore the CENVAR software was used to tabulate the standard errors for key survey estimates from the 1997/98 PHS data. This software package and the formulas used in the variance estimation procedures are described in Section 5.3.

The CENVAR results are presented in Annex III. For each estimate, the CENVAR table shows the standard error, coefficient of variation (CV), 95 percent confidence interval and design effect. The CV, or relative standard error, is a measure of the relative precision of each survey estimate, so it is a useful estimator for comparing the reliability of different survey estimates of totals and averages. In the case of the survey estimates of total crop area and production, the CVs vary mostly by the number of sample households growing the particular crop. The table for the first set of estimates in Annex III shows the CENVAR results for total production by crop (in metric tons). It can be seen in this table that 7 out of the 16 crops have CVs greater than 20 percent for total crop production, indicating a relatively low reliability for these estimates. The highest CV is for soybeans (50.8 percent), followed by Virginia tobacco (44.5 percent), Burley tobacco (43.5 percent) and Irish potatoes (40.9 percent); the number of sample households growing these crops are 82, 21, 29 and 31, respectively. On the other hand, maize is grown by 4,090 sample households, so the CV for the survey estimate of total maize production is only 4.2 percent.

The CENVAR results illustrate that the main limitation of the current sample design is that it does not provide reliable results for rare crops. In order to improve the reliability of survey estimates for these crops, it would be necessary to increase the number of sample households growing these crops. At the same time, given the limited resources for conducting the survey, it is not feasible to increase the overall sample size very much. Therefore a different sampling strategy would be required for increasing the number of sample households with rare crops. This is discussed in Section 4.5.

4. Recommendations for Improving Sample Design for Future PHS

4.1. Universe for PHS - Definition of Farm Households

The current sampling frame for the PHS includes all households in the rural areas of Zambia. A listing of all households is conducted within each sample SEA, and the sample households are selected from this listing (separately for Category A and B households). Some of the selected households are involved mainly in non-agricultural economic activities, although they may have a small house garden or a few poultry for household consumption. One problem with including non-farm households in the survey is that it reduces the effective sample size for the analysis of agricultural and livestock characteristics. This is especially important since the current sample size is relatively small given the comprehensive set of variables being measured. After examining the high sampling errors for some of the crops which are rare or limited to particular areas, it has been determined that it will be necessary to increase the number of sample households growing these crops. The non-farm households in the sample also confound the analysis for survey estimates such as average farm size, so it is sometimes necessary to filter out such households for the analysis.

Given that the PHS is primarily an agricultural survey, it is recommended that the universe and corresponding sampling frame be limited to farm households. The CSO can examine previous PHS results and consult with the main data users in determining the most appropriate definition for a farm household. In some developing countries the definition of a farm household is defined in terms of a minimum farm size or number of animals (with a different minimum for each type of animal).

For this reason it is recommended to first screen out non-farm households at the listing stage when possible, and eliminate them from the selection of in-scope farm households. The listing sheet should have a screening question with a clear definition for farm household, in order to identify the non-farm households which should be excluded from the selection at the last sampling stage.

It will also be necessary to screen out non-farm households in the PHS questionnaire itself, given possible changes between the time of the listing and the survey. Item 12 on the front page of the 1997/98 PHS questionnaire has a check list for crop growing (12.1), livestock raising (12.2), poultry raising (12.3) and fish farming (12.4). This check list should be modified to correspond to the minimum criteria for farm households (in terms of farm size or animals). There should be a skip pattern so that the interview would be ended for any household which does not meet the minimum criteria for agricultural activities (that is, a non-farm household).

4.2. Replacing Sample SEAs Not Currently Covered in PHS

For the 1996/97 PHS, 22 out of the 405 sample SEAs were not covered by the survey because of inaccessibility or security reasons. Although the number of missing sample SEAs may have decreased slightly for the 1997/98 PHS, it is still a problem. For example, some of the missing sample SEAs were found to be in swampy areas which were difficult to reach. One problem with such missing sample SEAs is that they can have a more significant effect on the bias of the survey results, as well as reduce the effective number of sample SEAs and households in the survey data. Given that the SEAs were selected systematically (with PPS) within each district, a missing sample SEA means that a part of the district is not represented in the survey. This is especially important for districts with only a few sample SEAs. In the case of one district (Chavuma in Northwestern Province), one of the two sample SEAs was not covered, so that half of the district is not represented in the survey estimates. Although this may only have a small effect on the national-level estimates, the provincial-level estimates would be more affected. In the case of this district with survey data for only one sample SEA, it was necessary to "collapse" (combine) this district with another one within the province for calculating the standard errors with CENVAR.

In order to reduce this source of bias and maintain the effective sample size, the CSO may consider selecting a replacement sample SEA for each one which cannot be covered by the survey. Sometimes it may be possible to select a new sample SEA within the same sample CSA, although in most cases the entire sample CSA may be inaccessible. It is important to use sampling procedures similar to those used for selecting the original sample SEAs in selecting the replacement sample SEAs. Also, each new sample SEA should be selected from the same part of the district as the original sample SEA which it is replacing. Although some of the CSAs in this part of the frame may also be inaccessible, the replacement SEA should be as close as possible to the original sample SEA.

One procedure which can be used to select the replacement sample SEAs would be to check the original systematic selection of CSAs in the frame, and identify all the accessible CSAs in the district within one sampling interval around the missing sample CSA; sometimes it may be necessary to go beyond one sampling interval to identify at least 4 eligible CSAs. These accessible CSAs near the missing sample CSA should be listed with the corresponding number of households in the frame, and the measures of size should be cumulated in order to select one new sample CSA with PPS. Then the replacement sample SEA would be selected with PPS within this new CSA. Perhaps this replacement procedure could be tested starting with the missing sample SEA in Chavuma District, which currently has only one sample SEA covered in the survey.

4.3. Stratification by Farm Size

The households with a larger farm size have a greater relative contribution to the survey estimates of total crop area and production and the corresponding variances. Therefore one way to increase the efficiency of the sample design for producing reliable estimates of total crop area and production and average crop yield would be to improve the farm size stratification for the households listed within sample SEAs. The current farm size stratification of households listed in the sample SEAs has two categories: A - households with less than 5 ha.; and B - households with 5 or more ha. The Category B households are included in the sample at the last stage with certainty or with a high probability. This stratification is helping to reduce the variances some, but it can be improved. One problem with the current classification of the listed households by category is the difference between the listing and survey information on farm size. Table 2 shows the distribution of the sample households from the 1997/98 PHS by listing category and farm size reported in the survey. It can be seen in this table that more than half of the Category B farms are found to have less than 5 ha. in the survey data. However, most of the misclassified Category B households report more than 2 ha. in the survey data, so this sampling approach is still effective in increasing the sample size for the larger farms.

The first recommendation is that the CSO try to improve the quality of the farm size information from the listing. It will never be perfect, but even a small reduction in the farm size misclassification will improve the sampling efficiency.

Table 2. Distribution of Sample Households from 1996/97 Post-Harvest Survey by Farm Size Reported in Survey and Category from Listing

Farm Size Group	Number of Sample Households		
	Total	Category A	Category B
Total	6,154	5,559	595
0 ha.	120	117	3
0.01-0.99 ha.	2,639	2,614	25
1.00-1.99 has.	1,894	1,819	75
2.00-2.99 has.	695	623	72
3.00-3.99 has.	297	224	73
4.00-4.99 has.	160	89	71
5.00-9.99 has.	303	66	237
10+ has.	46	7	39

The second recommendation is to improve the effectiveness of the farm size stratification. The current farm size cut-off for Category B is 5 ha. Given the small number of households in this category, less than 10 percent of the sample households are in Category B. There were only 595 sample households in Category B in all of the 405 sample PSUs selected for the 1997/98 PHS, so the average number of Category B households per sample PSU is less than 2. This limits the effectiveness of the farm size stratification.

Table 3 shows the weighted distribution of the households and area planted under crops by farm size group. Based a review of this distribution, it was found that 2 ha. would be an effective cut-off for the medium-scale farms. It can be seen in this table that 19.6 percent of the households have a farm size of 2 or more ha., which cover about 47.8 percent of the total farm area. However, the 5 ha. cut-off can still be used to stratify the largest farms; the households with 5 or more has. represent 2.1 percent of the total households but 11.3 percent of the total farm area. Therefore it is recommended to stratify the households from the listing into three farm size categories: 0-1.99 has., 2.00-4.99 has. and 5+ has. The households with 5 or more has. should continue to be selected with certainty at the last sampling stage (up to 10 sample households) within each sample SEA. The sample allocation for the other two farm size strata is described in Section 4.5.

The CENVAR software was used to tabulate the standard errors of total crop production by category and farm size group, to examine the effect of the farm size stratification on the standard errors of the estimates of total crop production. These CENVAR results are included in Annex III. In order to illustrate the variance components by farm size stratum, Table 4 shows the variances for total production of maize, which is one of the most predominant crops. For each estimate, the variance is the square of the standard error tabulated by CENVAR.

Table 3. Weighted Distribution of Total Households by Farm Size from 1997/98 Post-Harvest Survey Data

Farm Size Group	Weighted Total No. Households	Percent Households	Weighted Total Farm Area (Ha.)	Percent Farm Area
Total	897,149	100.0%	1,187,811	100.0%
0 ha.	19,396	2.2%	0	0.0%
0.01-0.99 ha.	410,342	45.7%	230,418	19.4%
1.00-1.99 has.	291,526	32.5%	389,700	32.8%
2.00-2.99 has.	101,542	11.3%	236,241	19.9%
3.00-3.99 has.	38,589	4.3%	126,198	10.6%
4.00-4.99 has.	16,859	1.9%	71,555	6.0%
5.00-9.99 has.	16,833	1.9%	104,620	8.8%
10+ has.	2,062	0.2%	29,079	2.4%

Table 4. Survey Estimates of Total Production of Maize (in Metric Tons) by Farm Size, and Corresponding Variance Components

Farm Size Group	Total Production (Metric Tons)	Percent Production	Variance	Percent Variance	No. Sample Households
0.01-0.99 ha.	127,622	20.5%	47,196,900	9.3%	1,570
1.00-1.99 has.	183,966	29.5%	98,386,561	19.4%	1,279
2.00-2.99 has.	107,638	17.3%	104,837,121	20.7%	524
3.00-3.99 has.	68,007	10.9%	77,281,681	15.2%	246
4.00-4.99 has.	43,485	7.0%	42,797,764	8.4%	139
5.00-9.99 has.	69,167	11.1%	76,930,441	15.2%	286
10+ has.	23,696	3.8%	60,202,081	11.9%	46
Total	623,581	100.0%	507,632,549	100.0%	4,090

In Table 4 it can be seen that the households with 2 or more has. account for 50.0 percent of the estimate of total production of maize, and 71.4 percent of total variance. The households with 5 or more has. account for 14.9 percent of the total production of maize, and 27.1 percent of the total variance. This illustrates that the sampling efficiency will be increased by selecting the households with 2 or more has. with a higher probability than that for the households with less than 2 has.; the households with 5 or more has. should continue to be included with certainty at the last sampling stage.

4.4. Stratification for Special Crops

The measures of precision for total crop area and production shown in the CENVAR tables in Annex III indicate that the coefficients of variation (CVs) are very high for certain crops which are rare or limited to particular geographic areas. As indicated in Section 3, the particular crops which have the highest CVs for total production are soyabeans, Virginia tobacco, Burley tobacco and Irish potatoes; these crops also have the lowest number of observations in the sample. Three additional crops which may require special attention are ground beans, rice and sunflowers, which were grown by only 105, 157 and 164 households, respectively, in the 1997/98 PHS sample. In order to improve the precision for the survey estimates for these crops, it will be necessary to increase the number of sample households growing these crops.

There are two ways to increase the number of sample households growing specific crops. One method would be to select the households growing these crops in the current sample SEAs with certainty at the last sampling stage (similar to the current procedure for Category B households). This would require an additional screening question in the listing sheet to identify the households growing these specific crops. Given that this method would be the easiest to implement, it is recommended to start using it as soon as possible. The effectiveness of this approach depends on the dispersion of the special crops among the sample SEAs. This can be studied using the crop data from the 1990 census for the sample SEAs. It is also useful to examine the number of sample households growing the special crops by SEA in the current sample. The 1997/98 PHS data were used to tabulate the number of sample SEAs by the number of sample households growing particular rare crops. These results are shown in Table 5.

It can be seen in Table 5 that most of the SEAs have only one or two sample households growing these particular crops, although there are a few SEAs where certain crops are concentrated. This distribution also indicates that there are probably other sample SEAs with households growing the specified crops, but none of these households were selected. Therefore the approach of selecting the households growing these crops with certainty at the last sampling stage should also increase the number of sample SEAs with observations for these crops.

In sample SEAs with many households growing a particular crop, it would not be efficient to include all of them in the sample with certainty at the last sampling stage because of the

Table 5. Distribution of Sample SEAs by Number of Sample Households Growing Particular Crops, from 1997/98 PHS Data

Crop	No. Sample Hhs.	Number of Sample SEAs with Crop										
		Total	Number of Sample Households Growing Crop in SEA									
			1	2	3	4	5	6	7	8	9	10+
Soyabeans	82	34	21	9			1					3
Irish Potatoes	31	16	11	2			2	1				
Virginia Tobacco	21	10	5	1	3		1					
Burley Tobacco	29	9	4	1	2			1				1
Ground Beans	105	55	33	10	6	2	1	1	1	1		
Cowpeas	44	25	15	6	1	1	2					

clustering effect. Therefore there should be a limit on the number of households with each crop to be included in the sample within each SEA. One option would be to select all the households growing a particular crop with certainty if there are four or less of these households in the sample SEA. Otherwise, these households could be selected within their respective farm size categories.

Table 5 also shows the total number of sample SEAs in which at least one sample household grew a particular crop. This illustrates the importance of the geographical dispersion of the crops. For example, it can be seen that Burley tobacco is found in only 9 sample SEAs. If such crops are only grown in particular parts of the country for various reasons such as soil and climatic conditions, the procedure of including the households growing these crops in the sample SEAs with certainty at the last sampling stage would be less effective for increasing the number of observations for these crops. In this case, most the sample SEAs would not have any households growing such crops. The only way to increase the number of observations for these localized crops would be to select additional sample SEAs where these crops are grown; this is the second method for improving the level of precision for the special crops.

The second approach would require a new stratification of the sampling frame to identify the SEAs in which particular crops are grown. The SEAs within each district would have to be divided into two substrata: those with special crops and those without. Additional sample SEAs could then be selected within the stratum with special crops (excluding the SEAs in the current sample), while keeping the current sample of SEAs. For estimation purposes, it would be necessary to divide the current sample SEAs into the two strata within each province. The weights for the sample households in the stratum with special crops would have to be adjusted based on the new number of sample SEAs in this stratum.

Of course, the second approach would require additional resources for listing and conducting the survey since the sample size would be increased. If resources are limited, the CSO may consider implementing the first method initially to determine its effectiveness for improving the level of precision of the survey estimates for the special crops. The second method would only be implemented once the required resources become available.

If the CSO decides to increase the number of sample SEAs with the special crops, the first method of including households growing special crops with certainty within sample SEAs should be also be implemented. That is, the second procedure should include the first method.

Following the next agricultural census, it is recommended to develop a new sampling frame with special stratification for the important rare or localized crops.

4.5. Sample Size and Allocation

The sample size for a particular survey is determined by the accuracy required for the survey estimates for each domain, as well as by the resource and operational constraints. The accuracy of the survey results depends on both the sampling error, which can be measured through variance estimation, and the nonsampling error, which can only partially be measured through expensive reinterview or validation studies. The sampling error is inversely proportional to the sample size. On the other hand, the nonsampling error may increase with the sample size, since it is more difficult to control the quality of a larger operation. It is therefore important that the overall sample size be manageable for quality and operational control purposes.

The sample size and allocation for the 1996/97 PHS is shown in Table 1. The sample size for this survey largely depends on the available resources. As indicated in previous sections, the current level of precision of the survey estimates for particular crops is not satisfactory. Some of the recommended procedures for increasing the number of observations for these crops would require additional resources.

The CENVAR results on standard errors for selected survey estimates could be used to determine the approximate sample size which would be required to increase the precision of the estimates for certain crops to a specified level. However, even if the current sample size were doubled, the CVs for certain rare crops may still not be satisfactory. For example, if the number of sample households with Burley tobacco were doubled, there would still only be about 60 observations for this crop. Therefore it is recommended that the CSO and the Ministry of Agriculture, Food and Fisheries (MAFF) prioritize the objectives of the PHS for next year, so that an efficient sampling frame can be developed based on the 2000 Census results. The CSO also needs to develop a strategic plan for its overall household survey program.

The current proportional allocation of the sample among the provinces is efficient for national-level estimates. However, if the provincial-level estimates become more important over time, the CSO may consider increasing the sample size for some of the smaller provinces such as Lusaka and Copperbelt. For example, a minimum sample of 30 SEAs for such provinces would improve the level of precision of the survey results for these domains.

One of the most important aspects of the sample design which affects the reliability of survey estimates is the number of sample SEAs and the number of households selected per sample SEA. The sampling efficiency is measured by the design effect, which is defined as the variance of a survey estimate based on a particular sample design (such as the stratified multi-stage sample design for the PHS) divided by the corresponding variance based on a simple random sample of the same size. In the case of a multi-stage sample, most of the design effect is due to the clustering effect from the homogeneity of household characteristics within a PSU (measured by the intraclass correlation). In the case of an agricultural survey such as the PHS, the similarity of cropping patterns for households with an SEA will increase the design effect.

It can be seen in the CENVAR tables in Annex II that 7 of the 16 crops have a design effect (DEFF) greater than 3 for total crop production. This indicates a relatively high clustering effect. One way to increase the sampling efficiency would be to select more sample SEAs and reduce the number of households selected per PSU. For example, a sample of 540 SEAs and 15 sample households per SEA would result in a total sample of 8,100 households. This is the same as the number of sample households selected under the current design of 405 sample SEAs and 20 sample households per SEA. However, the sample of 540 SEAs would increase the dispersion of the sample and increase the sampling efficiency.

Of course, increasing the number of sample SEAs will also increase the cost of the survey, given the proportionately higher cost of listing and the additional transportation costs in traveling between sample SEAs during the survey. If it is not possible to increase the number of sample SEAs, the current sampling strategy of selecting 20 households per sample SEA can be maintained.

In the sample design for the PHS, another important issue is the allocation of the 20 sample households within each sample SEA among the farm size categories (and special crops, if the procedures recommended in Section 4.4 are implemented). In Section 4.3 it was recommended to use the following three farm size categories:

Category A - 0 - 1.99 has.
Category B - 2.00 - 4.99 has.
Category C - 5+ has.

Based on this definition, the new Category C corresponds to the current Category B (based on two farm size categories). In order to specify the recommendations, the following terms are defined:

N = total number of households listed in the sample SEA
N_A = number of households listed in category A within the sample SEA
N_B = number of households listed in category B within the sample SEA
N_C = number of households listed in category C within the sample SEA
n_A = number of sample households selected in category A within the sample SEA
n_B = number of sample households selected in category B within the sample SEA
n_C = number of sample households selected in category C within the sample SEA

The following steps are recommended to allocate the 20 sample households by category within each sample SEA:

(1) If N_C is less than or equal to 10, select all the N_C households in Category C with certainty at the last sampling stage (that is, $n_C = N_C$).

(2) If N_B is less than or equal to 10-N_C, include all of the Category B households with certainty (in this case, $n_B = N_B$).

(3) If N_B is greater than 10-N_C, select 10-N_C households (systematically with a random start) from the Category B households in the listing (that is, $n_B = 10\text{-}N_C$).

(4) In the case of Category A, the number of households to be selected would determined as follows:
$n_A = 20\text{-}N_B\text{-}N_C$, if ($N_B$+$N_C$) is less than 10;
$n_A = 10$, if $n_B + n_C = 10$

In the current sample for the PHS, there are a few sample SEAs with more than 10 households with 5 or more has. (that is, in Category C). In order to simplify the sample allocation procedures for such SEAs, it is recommended to select 10 households in Category

C, 5 households in Category B and 5 households in Category A. The larger number of households in the listing for Category A will generally ensure that the sampling rate for Category B is higher than that for Category A.

The proportion of households in each farm size category will vary by sample SEA. In an SEA with a large number of households in Category B, it is possible that these procedures could result in the sampling rate for Category B being less than that for Category A. It is recommended that the sampling rate for the Category B households in each sample SEA be at least twice that for Category A. Therefore after determining the values of n_A, n_B and n_C from the four steps described above, it will be necessary to calculate the last stage sampling rates for Category A and B as follows:

$$f_A = \frac{n_A}{N_A}$$

$$f_B = \frac{n_B}{N_B}$$

If f_B is less than 2 times f_A, it is recommended to adjust the number of households to be selected in each category in order to satisfy the following equation:

$$n_A^{/} + n_B^{/} + n_C = n_A^{/} + 2 \times \frac{n_A^{/}}{N_A} \times N_b + n_C = 20 \ ,$$

where:

n'_A = adjusted number of sample households to be selected in Category A within the sample SEA

n'_B = adjusted number of sample households to be selected in Category B within the sample SEA

Solving this equation for n'_A results in the following expression:

$$n_A^{/} = \frac{N_A \times (20 - n_C)}{N_A + (2 \times N_B)}$$

$$n_B^{/} = 20 - n_C - n_A^{/}$$

As an example, assume that N_A = 50, N_B = 25 and N_C = 5 in the listing for a particular sample SEA. Solving these equations, the following values would be calculated:

$n'_A = 7.5$
$n'_B = 7.5$

In this case, the adjusted number of households to be selected in Category B (n'_B) should be rounded up to 8 and that for Category A (n'_A) should be rounded down to 7, so that the sampling rate for Category B is at least twice that for Category A.

Given that only about 20 percent of the households have more than 2 has., there should only be a few sample SEAs where it will be necessary to use this adjustment for the number of sample households to be selected in Categories A and B. If the sample selection is conducted in the field and the data collection staff would find it difficult to implement this refinement in the allocation procedures, the instructions could be limited to steps (1) to (4) described previously. If the selection of sample households is conducted in the office, it will be easier to control the improved allocation scheme.

If the CSO decides to implement the special procedures for selecting more households growing rare crops, it will also be necessary to identify these households in the listing. The procedures used to select the households with these crops will depend on the number of such households identified in the listing for the sample SEA. If the total number of sample households to be selected in each sample SEA is limited to 20, the households with rare crops can be assigned to Category B or C, regardless of their farm size, in order to simplify the sample selection procedures. Then the methodology described previously can be used to determine the number of sample households to be selected in each category. If the number of households with a particular rare crop in the SEA is four or less, they can be assigned to Category C, in order to be included in the sample with certainty at the last sampling stage. If there are more than four households with the rare crop, they can be assigned to Category B in order to be selected either with certainty or with a higher sampling rate.

Given the high CVs for the PHS estimates of total number of livestock by province shown in Annex III, the CSO may also consider identifying any households with a large number of livestock in the listing, to be included with certainty at the last sampling stage (for example, by including them with the Category C households). In the 1997/98 PHS data there were 31 sample households with 50 or more head of cattle, and 7 sample households with 100 or more head of cattle. The household with the largest number of cattle (473 heads) reported a farm size of 2.84 ha., so it was selected with the Category A households, and has a higher weight. This is one source of the high variability in the survey data for livestock. The cut-off in the number of cattle for the households in the listing to be included with the Category C households should be studied further based on the sample distribution, but at least the households with 50 or more head of cattle should receive special treatment.

5. Estimation Procedures

5.1. Weighting Procedures

The CSO staff is currently using appropriate weighting procedures for the PHS. In order for the sample estimates from a particular survey to be representative of the population, it is necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample household would be equal to the inverse of its probability of selection (calculated by multiplying the probabilities at each sampling stage).

Based on the current sample design for the PHS, the probability of selection within each SEA is different for the households listed in Category A (with less than 5 has.) and those listed in Category B (with 5 or more has.). The probability of selection for sample households in each farm size category within a sample SEA can be generalized as follows:

$$p_{Shij} = \frac{m_h \times N_{hi}}{N_h} \times \frac{N_{hij}}{N_{hi}} \times \frac{n_{Shij}}{N_{Shij}} = \frac{m_h \times N_{hij} \times n_{Shij}}{N_h \times N_{Shij}} ,$$

where:

p_{Shij} = probability of selection for the sample households in farm size category S in the j-th sample SEA within the i-th sample CSA in stratum (district) h

m_h = number of sample PSUs selected in stratum h

N_h = total number of households in the frame for stratum h

N_{hi} = total number of households in the frame for the i-th sample CSA in stratum h

N_{hij} = total number of households in the frame for the j-th sample SEA within the
i-th sample CSA in stratum h

n_{Shij} = number of sample households selected in farm size category S from the listing for the j-th sample SEA within the i-th sample CSA in stratum h

N_{Shij} = total number of households in farm size category S from the listing for the j-th sample SEA within the i-th sample CSA in stratum h

The three terms in p_{Shij} correspond to the first, second and third stage probabilities of selection; at the first two sampling stages the CSAs and SEAs were selected with PPS, and at the third stage the households were selected with equal probability within each farm size category. It can be seen that p_{Shij} simplifies to the probability of selecting the SEA with PPS directly at the first stage.

Based on the current sampling procedures, in most sample SEAs the households in Category B are selected with certainty at the last sampling stage (that is, $n_{Shij} = N_{Shij}$), in which case these households have the same probability of selection as the sample SEA.

If the CSO implements the three farm size groups recommended in Section 4.3, the formula for the probability of selection will be the same, with different probabilities for each of the three farm size categories within the sample SEA.

The basic sampling weight is equal to the inverse of the probability of selection. Therefore the corresponding basic weight for the sample households in each farm size category would be calculated as follows:

$$W_{Shij} = \frac{N_h}{m_h \times N_{hij}} \times \frac{N_{Shij}}{n_{Shij}}$$

It should be noted that the sample households selected in each farm size category keep the specified weight, even if it is found later that the farm size was misclassified according to the survey data.

If the CSO decides to use the three farm size categories (A, B and C) recommended in Section 4.3, it would be necessary to calculate a separate basic weight for the sample households in each category within a sample SEA, using the specified formula.

It is also important to adjust the weights to take into account the noninterviews. It would be ideal to have a separate noninterview adjustment factor for the sample households within each farm size category in the sample SEA. The numerator of this adjustment factor would be the total number of valid households selected in the particular category within the sample SEA (excluding any vacant housing units or out-of-scope households); the denominator would be the number of completed household questionnaires. The final weight adjusted for noninterviews would be calculated as follows:

$$W'_{Shij} = W_{Shij} \times \frac{n'_{Shij}}{n''_{Shij}} ,$$

where:

W'_{Shij} = weight adjusted for noninterviews for the sample households in farm size category S in the j-th sample SEA within the i-th sample CSA in stratum h

n'_{Shij} = number of valid households selected in farm size category S in the j-th sample SEA within the i-th sample CSA in stratum h (excluding any vacant housing units or out-of-scope households)

n''_{Shij} = number of sample households in farm size category S with completed interviews in the j-th sample SEA within the i-th sample CSA in stratum h

The CSO has been implementing a similar procedure for adjusting the weights for noninterviews. Instead of first calculating the basic weight, the final weight is calculated directly by substituting the value of n_{Shij} with that for n''_{Shij} in the formula for the basic weight specified previously. In cases where $n'_{Shij} = n_{Shij}$ (that is, there are no out-of-scope households selected in the sample SEA), this procedure is equivalent to using the specified formula for the final adjusted weight. This is generally the case, since a recent listing is used for selecting the sample households for the PHS; therefore the current weighting procedures are reasonable. However, if some out-of-scope sample households are found in the sample SEAs, the noninterview adjustment procedures specified here are recommended. If the CSO decides to exclude non-farm households from the universe for the PHS analysis as described in Section 4.1, some sample households may be identified as out-of-scope at the time of the survey.

The CSO staff has also been calculating special weights for certain crops which are highly localized. In this case, they post-stratify the CSAs within each stratum (district) into two groups: those which have households with the specified crop in the frame, and those without. The weight for each localized crop is then modified as follows:

$$W_{CShij} = \frac{N_{Ch}}{m_{Ch} \times N_{hij}} \times \frac{N_{Shij}}{n_{Shij}} ,$$

where:

W_{CShij} = weight for localized crop C in sample households in farm size category S in the j-th sample SEA within the i-th sample CSA in stratum h

N_{Ch} = total number of households in the frame for the CSAs with crop C within stratum h

m_{Ch} = number of sample CSAs with crop C in the frame within stratum h

These special crop weights have not yet been implemented in the PHS tabulations. However, this is a reasonable approach when certain crops are only found in part of a district. This crop weighting procedure could be refined further by checking the crops in the frame at the SEA level, and redefining the terms N_{Ch} and m_{Ch} as follows:

N'_{Ch} = total number of households in the frame for the SEAs with crop C within stratum h

m'_{Ch} = number of sample SEAs with crop C in the frame within stratum h

5.2. Types of Survey Estimates

The most common survey estimates to be calculated from the household surveys are in the form of totals and ratios. The survey estimate of a total can be expressed as follows:

$$\hat{Y} = \sum_{h=1}^{L} \sum_{i=1}^{m_h} \sum_{k=1}^{20} W'_{hijk} \; y_{hijk} \quad ,$$

where:

W'_{hijk} = final weight for the k-th sample household in the j-th sample SEA within the i-th sample CSA in stratum h

L = number of strata

y_{hijk} = value of variable y for the k-th sample household in the j-th sample SEA within the i-th sample CSA in stratum h

The survey estimate of a ratio is defined as follows:

$$\hat{R} = \frac{\hat{Y}}{\hat{X}} \; ,$$

where $\hat{Y}$ and $\hat{X}$ are estimates of totals for variables y and x, respectively, calculated as specified previously.

In the case of multi-stage sampling, means and proportions are special types of ratios. In the case of the mean, the variable X, in the denominator of the ratio, is defined to equal 1 for each element so that the denominator is the sum of the weights. In the case of a proportion, the variable X in the denominator is also defined to equal 1 for all elements; the variable Y in the numerator is binomial and is defined to equal either 0 or 1, depending on the absence or presence, respectively, of a specified characteristic in the unit observed.

5.3. Ratio Estimation for Particular Crops

In the case of particular crops which have a high level of sampling error because they are rare or grown in limited geographic areas, it may be possible to improve the survey estimates through ratio estimation, assuming that independent data for the crop are available from other sources such as frames maintained by the MAFF or farming associations.

Ratio estimation involves the use of independent information for a survey variable such as area planted for a particular crop. For example, it can be used to estimate total crop production when the total area planted for the crop is known from another source. In this case, the average crop yield would be estimated from the survey data and then multiplied by the total area planted, as follows:

$$\hat{P}_C = \frac{(\sum_h \sum_i \sum_k W'_{hijk} \times y_{Chijk})}{(\sum_h \sum_i \sum_k W'_{hijk} \times x_{Chijk})} \times X_C \ ,$$

where:

y_{Chijk} = production of crop C for the k-th sample household in the j-th sample SEA within the i-th sample CSA in stratum h

x_{Chijk} = area planted for crop C for the k-th sample household in the j-th sample SEA within the i-th sample CSA in stratum h

X_C = good estimate of total area planted in crop C from independent source

The first term represents the survey estimate of the average crop yield per hectare. Of course, one limitation of this ratio estimation procedure is the availability of accurate information on the total area planted for the particular crop. However, such data may be available for particular crops such as tobacco which may have farmer associations or special arrangements with a factory.

In other cases such as cotton, an accurate figure for crop production may be available from a processing or marketing company. In this case the total production of cotton from the independent source can be divided by the survey estimate of the average yield for cotton in order to estimate the total area planted in cotton.

5.4. Calculation of Variances

In the publication of the results from each survey it is important to include a statement on the accuracy of the survey data. In addition to presenting tables with calculated sampling errors for the most important survey estimates, the different sources of nonsampling error should be described.

The standard error, or square root of the variance, is used to measure the sampling error, although it may also include a small part of the nonsampling error. The variance estimator should take into account the different aspects of the sample design, such as the stratification and clustering. In order to avoid the time and effort it would require to develop custom variance programs, it is ideal to use an available software package to tabulate the variances. One such program available for calculating the variances for survey data from stratified multi-stage sample designs such as the master sample is CENVAR, a component of the Integrated Microcomputer Processing System (IMPS). CENVAR is menu-driven and user-friendly. It uses the data dictionary defined in the DATADICT component of IMPS. It can be used to calculate the variances of totals, means, proportions and other ratios. It produces subpopulation estimates for each category of a classification variable, and these variables can be cross-classified. For each estimate, CENVAR calculates the standard error, coefficient of variation (CV), 95 percent confidence interval and the design effect (DEFF). This software package uses an ultimate cluster variance estimator.

More than 10 years ago some CSO staff had experience in using the software package PCCARP, which is the predecessor of CENVAR. Only a DOS version of CENVAR is currently available, but it is much more user-friendly than PCCARP. Since the CSO has been using the IMPS software for processing the survey data, they have a copy of the CENVAR software and manual.

In order to tabulate estimates of standard errors using CENVAR, it is generally necessary to produce a new data input file from the original survey data. Since the CENVAR package will only accept one record type, it is necessary to generate one record for each unit of analysis in the CENVAR data input file. For example, in the case of the estimates by household, such as the average farm size per household, the CENVAR input file should have one record for each in-scope sample household. Each record in the CENVAR data input file should include fields for the stratum, cluster and weight, in addition to the classification and analysis variables which are required for the particular CENVAR analyses. The classification variables are used to produce subpopulation estimates for all their respective categories. The analysis variables are generally continuous variables, such as crop area and production, or count variables, which are equal to 1 if the unit has a certain characteristic and 0 otherwise. CENVAR automatically creates a count variable named INTERCEPT, which is equal to 1 for each record. The INTERCEPT variable can be used to obtain the estimate of the weighted total number of units (for example, the total number of households), or it can be used in the denominator of a ratio in order to obtain a mean or proportion; it can also be used as a classification variable to obtain estimates at the national level.

CENVAR does not accept any blanks in the file. In the case of classification variables, any record with a blank should be imputed with a special code to identify "missing" or "not applicable." The CENVAR output will include estimates for these categories, which can be deleted from the tabulations which will be published. For analysis variables, CENVAR assumes that any missing values are imputed. Once the file is zero-filled, CENVAR will treat any missing value as 0, thus introducing a downward bias in the estimates of means when there are missing values. One way to resolve this problem is to generate an indicator variable for each variable which has missing values. This indicator variable would then be crossed with each classification variable in the subpopulation analyses in order to produce separate estimates for the records with valid data for that variable. The subpopulation estimates for the missing value categories can later be deleted from the CENVAR output tables. This procedure was used for the CENVAR application developed for the 1997/98 PHS data, described in Section 5.4.

The ultimate cluster variance estimator for a total used by CENVAR can be expressed as follows:

Variance Estimator of a Total

$$V(\hat{Y}) = \sum_{h=1}^{L} \left[\frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{Y}_{hi} - \frac{\hat{Y}_h}{n_h} \right)^2 \right],$$

where:

$$\hat{Y}_{hi} = \sum_{k=1}^{20} W_{hijk}^{/} \, y_{hijk}$$

$$\hat{Y}_h = \sum_{i=1}^{n_h} \hat{Y}_{hi}$$

The variance estimator of a ratio used by CENVAR can be expressed as follows:

Variance Estimator of a Ratio

$$V(\hat{R}) = \frac{1}{\hat{X}^2} \left[V(\hat{Y}) + \hat{R}^2 \, V(\hat{X}) - 2 \, \hat{R} \, COV(\hat{X},\hat{Y}) \right],$$

where:

$$COV(\hat{X},\hat{Y}) = \sum_{h=1}^{L} \left[\frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{X}_{hi} - \frac{\hat{X}_h}{n_h} \right) \left(\hat{Y}_{hi} - \frac{\hat{Y}_h}{n_h} \right) \right]$$

$V(\hat{Y})$ and $V(\hat{X})$ are calculated according to the formula for the variance of a total.

5.5. CENVAR Applications for 1997/98 PHS Data

Three different CENVAR applications were developed for tabulating standard errors for selected estimates from the 1997/98 PHS data. The tables with the CENVAR results are presented in Annex III. These estimates include total crop area and production, crop yield per hectare and average farm size, as well as estimates for livestock and chickens. These applications can be used as a prototype for producing additional CENVAR analysis with the survey data in the future.

The first step in developing each application involved identifying the survey estimates to be included in the CENVAR analyses. Then specifications were written to define the CENVAR data input file needed to produce these estimates. It was necessary to generate a new text file with the specific data and recodes required for all the CENVAR analyses. This file contains one record for each sample household, given that the CENVAR software only accepts one record type. The specifications for the CENVAR data input file are presented in Annex I. This CENVAR file was generated from the original data file using an SPSS program, but other software such as the CONCOR component of IMPS, or a database program, can be used to produce this file.

Given that the CENVAR analyses pertained to characteristics of farm households, the records for any non-farm households in the original data file were excluded from the CENVAR file.

The first variables to be specified for the CENVAR data input file are those related to the sample design. In CENVAR it is necessary to identify the strata and the clusters (PSUs). For the 1997/98 PHS, the strata correspond to the provinces, and each SEA is defined as a cluster, since one sample SEA was selected within each sample CSA. In order to have a unique identification number for each sample SEA, the following 8-digit code was used to define the CLUSTER in the CENVAR file:

CLUSTER = DISTRICT (3 digits) + CSA (3 digits) + SEA1 (1 digit) + SEA2 (1 digit)

The PHS sample is stratified by district. Given that the district code is the first three digits of the new CLUSTER code, the stratum code can be identified as a subitem of the CLUSTER field in the IMPS data dictionary.

The variance formulas used by CENVAR require at least two sample clusters per stratum. However, because of a missing sample SEA in Chavuma District (code 701) in Northwestern Province, this stratum has only one sample SEA. Therefore it was necessary to "collapse" (combine) this stratum with a neighboring district, Zambesi (code 707), within the same province. This was carried out by simply changing each district code 701 in the CENVAR data file to 707, and then resorting the file by CLUSTER code. If the DISTRICT field were used as a classification variable in a CENVAR analysis, it would be necessary to generate a new variable with the correct district code in the CENVAR data file, or include a footnote in the table explaining that Chavuma District was combined with Zambesi District. However, given the small number of sample households at the district level, it is not recommended to tabulate the survey data at this level.

The third sampling variable in the CENVAR data file is the WEIGHT, corresponding to the final weight for each sample household used in producing the survey tables. In this application, the weight has four decimal places. The decimal point should not appear in the IMPS data files; the dictionary defines the number of implied decimals. For this reason the specifications indicate that the weight should be multiplied by 10,000.

Following the sampling variables, the data file should include all the classification variables for the CENVAR analyses by subpopulation. The names and codes of the categories for each classification variable should be defined in the IMPS data dictionary, and CENVAR will tabulate the estimates for each category of these variables. Since the province code is the first digit of the district code, the PROVINCE classification variable is also identified as a subitem of the CLUSTER field in the IMPS data dictionary. The second classification variable included in the CENVAR file is CATEGORY, which corresponds to the farm size stratum in which the sample household was selected from the listing; this variable was included in the data file as part of the evaluation of the current sample design. A third classification variable, farm size group, was generated from the farm size variable in the survey data. It was necessary to generate new codes for the farm size groups, since it is not possible to specify a range of values for classification variables in CENVAR.

Finally, the data file includes all of the analysis variables included in the CENVAR analyses. These are mostly continuous variables, such as total farm size, area planted and production for each crop, and number of animals. The following crops were included: maize, sorghum, rice, millet, sunflower, groundnuts, soybeans, seed cotton, Irish potatoes, Virginia tobacco, Burley tobacco, mixed beans, ground beans, cowpeas, sweet potatoes and cassava. In the case of crop production, the original data were reported in different units of measure for the various crops. In order to standardize these units, the value of production for each crop was converted to kilograms (kgs.) This required a transformation of the original data in generating the CENVAR input file. The data file also includes variables for the total number of cattle and chickens.

For each variable with decimal places in the original data, it was necessary to multiply the value by 10^x, where x is the number of decimal places, and then specify the number of implied decimal places in the IMPS dictionary.

In order to obtain the correct averages for farm size, cattle and chickens, the CENVAR data file includes indicator variables for each. In the case of farm size, the variable VALIDFS is equal to 1 if the household has a valid entry for farm size, and 0 if there is a missing value for farm size. By using this indicator as the denominator for mean farm size, the average is calculated only across households with valid farm size data. In the case of cattle and chickens, the indicator variables (HHCATTLE and HHCHICK) were equal to 1 if the household raised the respective animals, and 0 otherwise. This made it possible to obtain the average number of cattle and chickens for households raising these animals.

For the CENVAR analyses of total crop production, it was necessary to convert the production values for each crop to metric tons. Since the IMPS data dictionary specifies the number of implied decimals, it was simply necessary to generate a new IMPS dictionary for the CENVAR analyses of total crop production, in which the number of implied decimals for the production value of each crop was changed from 2 to 5. The IMPS data dictionary file CVPHS1A.DD specifies the crop production variables with two decimal places, while CVPHS1B.DD has five decimal places for crop production.

In the case of the variables for crop area and production, any record with missing data has a value of 0. Therefore the CENVAR estimates of total crop area and production represents the totals for households with valid data for the crop. When the missing values are not imputed, this will result in an underestimate of the true totals. The estimate of average yield per ha. for each crop is calculated as the total production divided by the total area for the particular crop.

Another CENVAR application was developed to study the average crop yield per household, by generating new crop yield variables for each sample household. A database program was used to generate the new CENVAR data file from the original file. For a particular crop, the yield variable in the sample household record was generated by dividing the crop production by the area planted (where the crop area was greater than 0). In the case of sample households with area planted in a particular crop but no production, the crop yield for that household is 0, indicating that the harvest was lost. It was also necessary to generate an indicator variable for each crop, which is equal to one if the household has valid area data for the crop, and 0 otherwise. This CENVAR application tabulated the average crop yield per household for the households growing the particular crop.

Annex II shows the listing for each of the three IMPS data dictionaries used for the CENVAR applications. The CLUSTER, PROVINCE and DISTRICT variables are defined in the COMMON part of each data dictionary, and the remaining variables are defined within the HOUSEHOLD record. It can be seen that in each data dictionary the PROVINCE and DISTRICT variables are defined as sub-items of the CLUSTER variable. Since there is only one record type, there is flexibility in terms of which variables are included in the COMMON part of the dictionary. However, it is only possible to specify implied decimals in the RECORD part of the dictionary, so the WEIGHT is included in the HOUSEHOLD record. The listing for each data dictionary also specifies the name and code of each category for the classification variables.

ANNEX I

Specifications for Generating CENVAR Input File for 1997/98 Post-Harvest Survey Data (Corresponding to IMPS Data Dictionaries CVPHS1A.DD and CVPHS1B.DD)

Specifications for Generating CENVAR Input File for 1997/98 Post-Harvest Survey Data (Corresponding to IMPS Data Dictionaries CVPHS1A.DD and CVPHS1B.DD)

In order to tabulate standard errors for the 1997/8 Zambia Post-Harvest Survey using CENVAR, a new ASCII (text) file was generated from the survey data with one record for each sample household. Each record had the following variables and format:

Name of Variable	Position	Specifications of Variable
DISTRICT	1-3	C002-DIST
CSA	4-6	C003-CSA
SEA1	7	C004-SEA1
SEA2	8	C005-SEA2
WEIGHT	9-16	C018-WGT (weight*10000)
CATEGORY	17	C017-CATEGORY
FSGROUP	18	FSGROUP = 0 for FL14-TOT-AREA = 0 = 1 for FL14-TOT-AREA = 0.01-0.99 = 2 for FL14-TOT-AREA = 1.00-1.99 = 3 for FL14-TOT-AREA = 2.00-2.99 = 4 for FL14-TOT-AREA = 3.00-3.99 = 5 for FL14-TOT-AREA = 4.00-4.99 = 6 for FL14-TOT-AREA = 5.00-9.99 = 7 for FL14-TOT-AREA = 10+
FARMSIZE	19-22	FL14-TOT-AREA*100 (that is, no decimals)
MAIZAREA	23-26	MAIZAREA = CL12-AREA-PLANT*100, if C000-CROP-CODE = 01 and CL12-AREA-PLANT is in valid range; MAIZAREA = 0 otherwise; including missing values and out-of-scope (households without maize)
MAIZPROD	27-34	MAIZPROD = CH148-PRODUCTION*9000 (that is, 100*kg), if C000-CROP-CODE = 01 and CH148-PRODUCTION is in valid range; MAIZPROD = 0 otherwise; including missing values and out-of-scope (households without maize)

Name of Variable	Position	Specifications of Variable
SORGAREA	35-38	SORGAREA = CL12-AREA-PLANT*100, if C000-CROP-CODE = 02 and CL12-AREA-PLANT is in valid range; SORGAREA = 0 otherwise; including missing values and out-of-scope (households without sorghum)
SORGPROD	39-46	SORGPROD = CH148-PRODUCTION*9000 (that is, 100*kg), if C000-CROP-CODE = 02 and CH148-PRODUCTION is in valid range; SORGPROD = 0 otherwise; including missing values and out-of-scope (households without sorghum)
RICEAREA	47-50	RICEAREA = CL12-AREA-PLANT*100, if C000-CROP-CODE = 03 and CL12-AREA-PLANT is in valid range; RICEAREA = 0 otherwise; including missing values and out-of-scope (farms without rice)
RICEPROD	51-58	RICEPROD = CH148-PRODUCTION*8000 (that is, 100*kg), if C000-CROP-CODE = 03 and CH148-PRODUCTION is in valid range; RICEPROD = 0 otherwise; including missing values and out-of-scope (farms without rice)
MILLAREA	59-62	MILLAREA = CL12-AREA-PLANT*100, if C000-CROP-CODE = 04 and CL12-AREA-PLANT is in valid range; MILLAREA = 0 otherwise; including missing values and out-of-scope (farms without millet)
MILLPROD	63-70	MILLPROD = CH148-PRODUCTION*9000 (that is, 100*kg), if C000-CROP-CODE = 04 and CH148-PRODUCTION is in valid range; MILLPROD = 0 otherwise; including missing values and out-of-scope (farms without millet)
SUNFAREA	71-74	SUNFAREA = CL12-AREA-PLANT*100, if C000-CROP-CODE = 05 and CL12-AREA-PLANT is in valid range; SUNFAREA = 0 otherwise; including missing values and out-of-scope (farms without sunflower)

Name of Variable	Position	Specifications of Variable
SUNFPROD	75-82	SUNFPROD = CH148-PRODUCTION*5000 (that is, 100*kg), if C000-CROP-CODE = 05 and CH148-PRODUCTION is in valid range; SUNFPROD = 0 otherwise; including missing values and out-of-scope (farms without sunflower)
GNUTAREA	83-86	GNUTAREA = CL12-AREA-PLANT*100, if C000-CROP-CODE = 06 and CL12-AREA-PLANT is in valid range; GNUTAREA = 0 otherwise; including missing values and out-of-scope (farms without groundnuts)
GNUTPROD	87-94	GNUTPROD = CH148-PRODUCTION*8000 (that is, 100*kg), if C000-CROP-CODE = 06 and CH148-PRODUCTION is in valid range; GNUTPROD = 0 otherwise; including missing values and out-of-scope (farms without groundnuts)
SOYAAREA	95-98	SOYAAREA = CL12-AREA-PLANT*100, if C000-CROP-CODE = 07 and CL12-AREA-PLANT is in valid range; SOYAAREA = 0 otherwise; including missing values and out-of-scope (farms without soyabeans)
SOYAPROD	99-106	SOYAPROD = CH148-PRODUCTION*9000 (that is, 100*kg), if C000-CROP-CODE = 07 and CH148-PRODUCTION is in valid range; SOYAPROD = 0 otherwise; including missing values and out-of-scope (farms without soyabeans)
COTTAREA	107-110	COTTAREA = CL12-AREA-PLANT*100, if C000-CROP-CODE = 08 and CL12-AREA-PLANT is in valid range; COTTAREA = 0 otherwise; including missing values and out-of-scope (farms without seed cotton)
COTTPROD	111-118	COTTPROD = CH148-PRODUCTION*100 (that is, 100*kg) , if C000-CROP-CODE = 08 and CH148-PRODUCTION is in valid range; COTTPROD = 0 otherwise; including missing values and out-of-scope (farms without seed cotton)

Name of Variable	Position	Specifications of Variable
POTAAREA	119-122	POTAAREA = CL12-AREA-PLANT*100, if C000-CROP-CODE = 09 and CL12-AREA-PLANT is in valid range; POTAAREA = 0 otherwise; including missing values and out-of-scope (farms without Irish potatoes)
POTAPROD	123-130	POTAPROD = CH148-PRODUCTION*1000 (that is, 100*kg), if C000-CROP-CODE = 09 and CH148-PRODUCTION is in valid range; POTAPROD = 0 otherwise; including missing values and out-of-scope (farms without Irish potatoes)
VTOBAREA	131-134	VTOBAREA = CL12-AREA-PLANT*100, if C000-CROP-CODE = 10 and CL12-AREA-PLANT is in valid range; VTOBAREA = 0 otherwise; including missing values and out-of-scope (farms without Virginia tobacco)
VTOBPROD	135-142	VTOBPROD = CH148-PRODUCTION*100 (that is, 100*kg), if C000-CROP-CODE = 10 and CH148-PRODUCTION is in valid range; VTOBPROD = 0 otherwise; including missing values and out-of-scope (farms without Virginia tobacco)
BTOBAREA	143-146	BTOBAREA = CL12-AREA-PLANT*100, if C000-CROP-CODE = 11 and CL12-AREA-PLANT is in valid range; BTOBAREA = 0 otherwise; including missing values and out-of-scope (farms without Burley tobacco)
BTOBPROD	147-154	BTOBPROD = CH148-PRODUCTION*100 (that is, 100*kg), if C000-CROP-CODE = 11 and CH148-PRODUCTION is in valid range; BTOBPROD = 0 otherwise; including missing values and out-of-scope (farms without Burley tobacco)
MBNSAREA	155-158	MBNSAREA = CL12-AREA-PLANT*100, if C000-CROP-CODE = 12 and CL12-AREA-PLANT is in valid range; MBNSAREA = 0 otherwise; including missing values and out-of-scope (farms without mixed beans)

Name of Variable	Position	Specifications of Variable
MBNSPROD	159-166	MBNSPROD = CH148-PRODUCTION*9000 (that is, 100*kg), if C000-CROP-CODE = 12 and CH148-PRODUCTION is in valid range; MBNSPROD = 0 otherwise; including missing values and out-of-scope (farms without mixed beans)
GBNSAREA	167-170	GBNSAREA = CL00000000000012-AREA-PLANT*100, if C000-CROP-CODE = 13 and CL12-AREA-PLANT is in valid range; GBNSAREA = 0 otherwise; including missing values and out-of-scope (farms without ground beans)
GBNSPROD	171-178	GBNSPROD = CH148-PRODUCTION*9000 (that is, 100*kg), if C000-CROP-CODE = 13 and CH148-PRODUCTION is in valid range; GBNSPROD = 0 otherwise; including missing values and out-of-scope (farms without ground beans)
CPEAAREA	179-182	CPEAAREA = CL12-AREA-PLANT*100, if C000-CROP-CODE = 14 and CL12-AREA-PLANT is in valid range; CPEAAREA = 0 otherwise; including missing values and out-of-scope (farms without cowpeas)
CPEAPROD	183-190	CPEAPROD = CH148-PRODUCTION*9000 (that is, 100*kg), if C000-CROP-CODE = 14 and CH148-PRODUCTION is in valid range; CPEAPROD = 0 otherwise; including missing values and out-of-scope (farms without cowpeas)
SPOTAREA	191-194	SPOTAREA = S172-AREA*100, if S171-GROW-SWEET = 1 and S172-AREA is in valid range; SPOTAREA = 0 otherwise; including missing values and out-of-scope (farms without sweet potatoes)
SPOTPROD	195-202	SPOTPROD = SH1740-PRODUCTIO*1000 (that is, 100*kg), if S171-GROW-SWEET = 1 and SH1740-PRODUCTIO is in valid range; SPOTPROD = 0 otherwise; including missing values and out-of-scope (farms without sweet potatoes)

Name of Variable	Position	Specifications of Variable
CASSAREA	203-206	CASSAREA = S172-AREA*100, if CL181-GROW-CASS = 1 and S172-AREA is in valid range; CASSAREA = 0 otherwise; including missing values and out-of-scope (farms without cassava)
CASSPROD	207-214	CASSPROD = CH1840-PRODUCTIO*9000 (that is, 100*kg), if CL181-GROW-CASS = 1 and CH1840-PRODUCTIO is in valid range; CASSPROD = 0 otherwise; including missing values and out-of-scope (farms without cassava)
HHCATTLE	215	HHCATTLE = 1 if OL11-RAISE-CATTL = 1 and OL13-NUMBER-TODA is in valid range; HHCATTLE = 0 otherwise; including missing values and records where OL11-RAISE-CATTL = 2
NOCATTLE	216-219	NOCATTLE = OL13-NUMBER-TODA if OL11-RAISE-CATTL = 1 and OL13-NUMBER-TODA is in valid range; NOCATTLE = 0 otherwise; including missing values and records where OL11-RAISE-CATTL = 2
HHCHICK	220	HHCHICK = 1if P00-POULTRY-CODE = 13 and P132-NO-TODAY is in valid range; HHCHICK = 0 otherwise; including missing values and out-of-scope (households with no chickens)
NOCHICK	221-224	NOCHICK = CA22-NUMBER-RAIS if P00-POULTRY-CODE = 13 and P132-NO-TODAY is in valid range; NOCHICK = 0 otherwise; including missing values and out-of-scope (households with no chickens)
VALIDFS	225	VALIDFS= 1if FL14-TOT-AREA is in valid range; VALIDFS = 0 otherwise (including missing values)

Notes:

All the values for the area of each crop are converted to ha.*100, since the dictionary for the CENVAR file will show two implied decimals. For the same reason, all the values for the production for each crop are converted to 100 kgs.

This file was sorted by CLUSTER, which includes the DISTRICT code. The file was zero-filled; that is, any blanks were replaced with 0.

ANNEX II

Listing of IMPS Data Dictionaries for CENVAR Applications Using 1997/98 Post-Harvest Survey Data

Listing of IMPS Data Dictionaries for CENVAR Applications Using 1997/98 Post-Harvest Survey Data

```
Page    1                Data Dictionary: CVPHS1A            IMPS Version 3.1
                         Created: 07/07/00  23:21:15

                         Record Length:      225

     The following records have been defined:

     Record               Record Type Value                          Max.
     Name                     (RECTYPE)          Required        Records
     ------               -----------------      --------        -------
     HOUSEHOLD                                      Y               1

The following COMMON items have been defined.  They occur on all records.
--------------------------------------------------------------------------------
Item (occurs)        Data            Item
  Subitem (occurs)   Type  Position  Len.      Value Name       Values
--------------------------------------------------------------------------------

CLUSTER                N       1-8      8

  PROVINCE             N       1        1      Central          1
                                               Copperbelt       2
                                               Eastern          3
                                               Luapula          4
                                               Lusaka           5
                                               Northern         6
                                               Northwestern     7
                                               Southern         8
                                               Western          9

  DISTRICT             N       1-3      3      Chibombo         101
                                               Kabwe Urban      102
                                               Kapiri Mposhi    103
                                               Mkushi           104
                                               Mumbwa           105
                                               Serenje          106
                                               Chililabambwe    201
                                               Chingola         202
                                               Kalulushi        203
                                               Kitwe            204
                                               Luanshya         205
                                               Lufwamyana       206
                                               Masaiti          207
                                               Mpongwe          208
                                               Mufulira         209
                                               Ndola Urban      210
                                               Chadiza          301
                                               Chama            302
                                               Chipata          303
                                               Katete           304
                                               Lundazi          305
                                               Mambwe           306
                                               Nyimba           307
                                               Petauke          308
                                               Chienge          401
                                               Kawambwa         402
                                               Mansa            403
Page    2                Data Dictionary: CVPHS1A            IMPS Version 3.1
                         Created: 07/07/00  23:21:15

The following COMMON items have been defined.  They occur on all records.
--------------------------------------------------------------------------------
Item (occurs)        Data            Item
  Subitem (occurs)   Type  Position  Len.      Value Name       Values
--------------------------------------------------------------------------------
```

Milenge	404
Mwense	405
Nchelenge	406
Samfya	407
Chongwe	501
Kafue	502
Luangwa	503
Lusaka Urban	504
Chilubi	601
Chinsali	602
Isoka	603
Kaputa	604
Kasama	605
Luwingu	606
Mbala	607
Mpika	608
Mporokoso	609
Mpulungu	610
Mungwi	611
Nakonde	612
Chavuma	701
Kabompo	702
Kasempa	703
Mufumbwe	704
Mwinilunga	705
Solwesi	706
Zambesi	707
Choma	801
Gwembe	802
Itezi-tezi	803
Kalomo	804
Kazungula	805
Livingstone	806
Mazabuka	807
Monze	808
Namwala	809
Siavonga	810
Sinazongwe	811
Kalabo	901
Kaoma	902
Lukulu	903
Mongu	904
Senanga	905
Sesheke	906
Shangombo	907

Record Name: HOUSEHOLD Record Type:

Item (occurs) Subitem (occurs)	Data Type	Position	Item Len.	Dec.	Value Name	Values
WEIGHT	N	9-16	8	4		
CATEGORY	N	17	1	0	Category A	1
					Category B	2
FSGROUP	N	18	1	0	0 ha.	0
					0.01-0.99 ha.	1
					1.00-1.99 has.	2
					2.00-2.99 has.	3
					3.00-3.99 has.	4
					4.00-4.99 has.	5
					5.00-9.99 has.	6
					10+ has.	7
FARMSIZE	N	19-22	4	2		
MAIZAREA	N	23-26	4	2		
MAIZPROD	N	27-34	8	2		
SORGAREA	N	35-38	4	2		
SORGPROD	N	39-46	8	2		
RICEAREA	N	47-50	4	2		
RICEPROD	N	51-58	8	2		
MILLAREA	N	59-62	4	2		
MILLPROD	N	63-70	8	2		
SUNFAREA	N	71-74	4	2		
SUNFPROD	N	75-82	8	2		
GNUTAREA	N	83-86	4	2		
GNUTPROD	N	87-94	8	2		
SOYAAREA	N	95-98	4	2		
SOYAPROD	N	99-106	8	2		
COTTAREA	N	107-110	4	2		
COTTPROD	N	111-118	8	2		
POTAAREA	N	119-122	4	2		

Record Name: HOUSEHOLD Record Type:

Item (occurs) Subitem (occurs)	Data Type	Position	Item Len.	Dec.	Value Name	Values
POTAPROD	N	123-130	8	2		
VTOBAREA	N	131-134	4	2		
VTOBPROD	N	135-142	8	2		
BTOBAREA	N	143-146	4	2		
BTOBPROD	N	147-154	8	2		
MBNSAREA	N	155-158	4	2		
MBNSPROD	N	159-166	8	2		
GBNSAREA	N	167-170	4	2		
GBNSPROD	N	171-178	8	2		
CPEAAREA	N	179-182	4	2		
CPEAPROD	N	183-190	8	2		
SPOTAREA	N	191-194	4	2		
SPOTPROD	N	195-202	8	2		
CASSAREA	N	203-206	4	2		
CASSPROD	N	207-214	8	2		
HHCATTLE	N	215	1	0		
NOCATTLE	N	216-219	4	0		
HHCHICK	N	220	1	0		
NOCHICK	N	221-224	4	0		
VALIDFS	N	225	1	0		

Record Length: 225

The following records have been defined:

Record Name	Record Type Value (RECTYPE)	Required	Max. Records
HOUSEHOLD		Y	1

The following COMMON items have been defined. They occur on all records.

Item (occurs) Subitem (occurs)	Data Type	Position	Item Len.	Value Name	Values
CLUSTER	N	1-8	8		
PROVINCE	N	1	1	Central	1
				Copperbelt	2
				Eastern	3
				Luapula	4
				Lusaka	5
				Northern	6
				Northwestern	7
				Southern	8
				Western	9
DISTRICT	N	1-3	3	Chibombo	101
				Kabwe Urban	102
				Kapiri Mposhi	103
				Mkushi	104
				Mumbwa	105
				Serenje	106
				Chililabambwe	201
				Chingola	202
				Kalulushi	203
				Kitwe	204
				Luanshya	205
				Lufwamyama	206
				Masaiti	207
				Mpongwe	208
				Mufulira	209
				Ndola Urban	210
				Chadiza	301
				Chama	302
				Chipata	303
				Katete	304
				Lundazi	305
				Mambwe	306
				Nyimba	307
				Petauke	308
				Chienge	401
				Kawambwa	402
				Mansa	403

The following COMMON items have been defined. They occur on all records.

Item (occurs) Subitem (occurs)	Data Type	Position	Item Len.	Value Name	Values
				Milenge	404
				Mwense	405
				Nchelenge	406
				Samfya	407
				Chongwe	501
				Kafue	502
				Luangwa	503
				Lusaka Urban	504
				Chilubi	601
				Chinsali	602
				Isoka	603
				Kaputa	604
				Kasama	605
				Luwingu	606
				Mbala	607
				Mpika	608
				Mporokoso	609
				Mpulungu	610
				Mungwi	611
				Nakonde	612
				Chavuma	701
				Kabompo	702
				Kasempa	703
				Mufumbwe	704
				Mwinilunga	705
				Solwesi	706
				Zambesi	707
				Choma	801
				Gwembe	802
				Itezi-tezi	803
				Kalomo	804
				Kazungula	805
				Livingstone	806
				Mazabuka	807
				Monze	808
				Namwala	809
				Siavonga	810
				Sinazongwe	811
				Kalabo	901
				Kaoma	902
				Lukulu	903
				Mongu	904
				Senanga	905
				Sesheke	906
				Shangombo	907

Record Name: HOUSEHOLD Record Type:

Item (occurs) Subitem (occurs)	Data Type	Position	Item Len.	Dec.	Value Name	Values
WEIGHT	N	9-16	8	4		
CATEGORY	N	17	1	0	Category A	1
					Category B	2
FSGROUP	N	18	1	0	0 ha.	0
					0.01-0.99 ha.	1
					1.00-1.99 has.	2
					2.00-2.99 has.	3
					3.00-3.99 has.	4
					4.00-4.99 has.	5
					5.00-9.99 has.	6
					10+ has.	7
FARMSIZE	N	19-22	4	2		
MAIZAREA	N	23-26	4	2		
MAIZPRMT	N	27-34	8	5		
SORGAREA	N	35-38	4	2		
SORGPRMT	N	39-46	8	5		
RICEAREA	N	47-50	4	2		
RICEPRMT	N	51-58	8	5		
MILLAREA	N	59-62	4	2		
MILLPRMT	N	63-70	8	5		
SUNFAREA	N	71-74	4	2		
SUNFPRMT	N	75-82	8	5		
GNUTAREA	N	83-86	4	2		
GNUTPRMT	N	87-94	8	5		
SOYAAREA	N	95-98	4	2		
SOYAPRMT	N	99-106	8	5		
COTTAREA	N	107-110	4	2		
COTTPRMT	N	111-118	8	5		
POTAAREA	N	119-122	4	2		

Record Name: HOUSEHOLD Record Type:

Item (occurs) Subitem (occurs)	Data Type	Position	Item Len.	Dec.	Value Name	Values
POTAPRMT	N	123-130	8	5		
VTOBAREA	N	131-134	4	2		
VTOBPRMT	N	135-142	8	5		
BTOBAREA	N	143-146	4	2		
BTOTPRMT	N	147-154	8	5		
MBNSAREA	N	155-158	4	2		
MBNSPRMT	N	159-166	8	5		
GBNSAREA	N	167-170	4	2		
GBNSPRMT	N	171-178	8	5		
CPEAAREA	N	179-182	4	2		
CPEAPRMT	N	183-190	8	5		
SPOTAREA	N	191-194	4	2		
SPOTPRMT	N	195-202	8	5		
CASSAREA	N	203-206	4	2		
CASSPRMT	N	207-214	8	5		
HHCATTLE	N	215	1	0		
NOCATTLE	N	216-219	4	0		
HHCHICK	N	220	1	0		
NOCHICK	N	221-224	4	0		
VALIDFS	N	225	1	0		

Created: 07/07/00 23:41:32

Record Length: 167

The following records have been defined:

Record Name	Record Type Value (RECTYPE)	Required	Max. Records
HOUSEHOLD		Y	1

The following COMMON items have been defined. They occur on all records.

Item (occurs) Subitem (occurs)	Data Type	Position	Item Len.	Value Name	Values
CLUSTER	N	1-8	8		
PROVINCE	N	1	1	Central	1
				Copperbelt	2
				Eastern	3
				Luapula	4
				Lusaka	5
				Northern	6
				Northwestern	7
				Southern	8
				Western	9
DISTRICT	N	1-3	3	Chibombo	101
				Kabwe Urban	102
				Kapiri Mposhi	103
				Mkushi	104
				Mumbwa	105
				Serenje	106
				Chililabambwe	201
				Chingola	202
				Kalulushi	203
				Kitwe	204
				Luanshya	205
				Lufwanyana	206
				Masaiti	207
				Mpongwe	208
				Mufulira	209
				Ndola Urban	210
				Chadiza	301
				Chama	302
				Chipata	303
				Katete	304
				Lundazi	305
				Mambwe	306
				Nyimba	307
				Petauke	308
				Chienge	401
				Kawambwa	402
				Mansa	403

The following COMMON items have been defined. They occur on all records.

Item (occurs) Subitem (occurs)	Data Type	Position	Item Len.	Value Name	Values
				Milenge	404
				Mwense	405
				Nchelenge	406
				Samfya	407
				Chongwe	501
				Kafue	502
				Luangwa	503
				Lusaka Urban	504
				Chilubi	601
				Chinsali	602
				Isoka	603
				Kaputa	604
				Kasama	605
				Luwingu	606
				Mbala	607
				Mpika	608
				Mporokoso	609
				Mpulungu	610
				Mungwi	611
				Nakonde	612
				Chavuma	701
				Kabompo	702
				Kasempa	703
				Mufumbwe	704
				Mwinilunga	705
				Solwesi	706
				Zambesi	707
				Choma	801
				Gwembe	802
				Itezi-tezi	803
				Kalomo	804
				Kazungula	805
				Livingstone	806
				Mazabuka	807
				Monze	808
				Namwala	809
				Siavonga	810
				Sinazongwe	811
				Kalabo	901
				Kaoma	902
				Lukulu	903
				Mongu	904
				Senanga	905
				Sesheke	906
				Shangombo	907

Record Name: HOUSEHOLD Record Type:

Item (occurs) Subitem (occurs)	Data Type	Position	Item Len.	Dec.	Value Name	Values
WEIGHT	N	9-16	8	4		
CATEGORY	N	17	1	0	Category A	1
					Category B	2
FSGROUP	N	18	1	0	0 ha.	0
					0.01-0.99 has.	1
					1.00-1.99 has.	2
					2.00-2.99 has.	3
					3.00-3.99 has.	4
					4.00-4.99 has.	5
					5.00-9.99 has.	6
					10+ has.	7
FARMSIZE	N	19-22	4	2		
MAIZYLD	N	23-30	8	2		
MAIZFARM	N	31	1	0	No maize	0
					Grows maize	1
SORGYLD	N	32-39	8	2		
SORGFARM	N	40	1	0	No sorghum	0
					Grows sorghum	1
RICEYLD	N	41-48	8	2		
RICEFARM	N	49	1	0	No rice	0
					Grows rice	1
MILLYLD	N	50-57	8	2		
MILLFARM	N	58	1	0	No millet	0
					Grows millet	1
SUNFYLD	N	59-66	8	2		
SUNFFARM	N	67	1	0	No sunflowers	0
					Grows sunflowers	1
GNUTYLD	N	68-75	8	2		
GNUTFARM	N	76	1	0	No groundnuts	0
					Grows groundnuts	1
SOYAYLD	N	77-84	8	2		

Record Name: HOUSEHOLD Record Type:

Item (occurs) Subitem (occurs)	Data Type	Position	Item Len.	Dec.	Value Name	Values
SOYAFARM	N	85	1	0	No soyabeans	0
					Grows soyabeans	1
COTTYLD	N	86-93	8	2		
COTTFARM	N	94	1	0	No cotton	0
					Grows cotton	1
POTAYLD	N	95-102	8	2		
POTAFARM	N	103	1	0	No potatos	0
					Grows potatoes	1
VTOBYLD	N	104-111	8	2		
VTOBFARM	N	112	1	0	No V. Tobacco	0
					Grows V. tabacco	1
BTOBYLD	N	113-120	8	2		
BTOBFARM	N	121	1	0	No B. tobacco	0
					Grows B. tobacco	1
MBNSYLD	N	122-129	8	2		
MBNSFARM	N	130	1	0	No mixed beans	0
					Grows mixed bean	1
GBNSYLD	N	131-138	8	2		
GBNSFARM	N	139	1	0	No ground beans	0
					Grows ground bns	1
CPEAYLD	N	140-147	8	2		
CPEAFARM	N	148	1	0	No cowpeas	0
					Grows cowpeas	1
SPOTYLD	N	149-156	8	2		
SPOTFARM	N	157	1	0	No sweet potato	0
					Grows sweet pot.	1
CASSYLD	N	158-165	8	2		
CASSFARM	N	166	1	0	No cassava	0
					Grows cassava	1
VALIDFS	N	167	1	0	No farm size	0
					Valid farm size	1

ANNEX III

Summary of CENVAR Tabulations of Standard Errors from Zambia 1997/98 Post Harvest Survey Data

Summary of CENVAR Tabulations of Standard Errors from Zambia 1997/98 Post Harvest Survey Data

1. Survey Estimate of Total Production in Metric Tons by Crop

Number of observations: 6154

Variable	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect
Maize	623,580	26,443	4.24	571,752	675,408	1.94
Sorghum	23,598	2,798	11.86	18,113	29,083	3.84
Rice	7,536	1,554	20.63	4,489	10,582	2.90
Millet	44,488	3,040	6.83	38,530	50,445	3.33
Sunflower	6,135	1,489	24.27	3,216	9,054	1.73
Groundnut	43,679	2,923	6.69	37,950	49,409	3.16
Soyabeans	2,371	1,204	50.77	12	4,731	5.39
Seed Cotton	70,933	11,472	16.17	48,448	93,417	5.27
Irish Potatoes	2,166	885	40.85	432	3,900	2.46
Virginia Tobacco	1,173	523	44.54	149	2,198	2.94
Burley Tobacco	3,493	1,520	43.52	513	6,473	3.55
Mixed Beans	9,912	1,308	13.20	7,347	12,476	2.99
Ground Beans	1,123	209	18.63	713	1,533	1.68
Cowpeas	486	151	30.97	191	782	1.66
Sweet Potatoes	38,413	3,852	10.03	30,864	45,963	2.73
Cassava	230,359	15,404	6.69	200,166	260,551	4.49

2. Survey Estimate of Total Area in Has. by Crop

Number of observations: 6154

Variable	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect
Maize	521,094	18,362	3.52	485,105	557,083	3.49
Sorghum	36,019	4,377	12.15	27,440	44,599	4.71
Rice	9,603	2,130	22.18	5,429	13,778	4.24
Millet	63,849	5,213	8.16	53,632	74,066	4.83
Sunflower	12,311	2,419	19.65	7,570	17,052	2.19
Groundnut	120,508	6,059	5.03	108,633	132,382	3.47
Soyabeans	4,902	1,673	34.12	1,623	8,180	4.14
Seed Cotton	79,305	10,250	12.93	59,214	99,395	5.58
Irish Potatoes	1,351	475	35.12	421	2,281	3.22
Virginia Tobacco	2,061	851	41.27	394	3,728	2.59
Burley Tobacco	3,464	1,665	48.07	200	6,729	4.76
Mixed Beans	24,009	2,598	10.82	18,917	29,100	2.05
Ground Beans	2,550	489	19.16	1,593	3,508	2.12
Cowpeas	2,266	631	27.84	1,029	3,503	2.53
Sweet Potatoes	22,216	2,631	11.84	17,059	27,372	1.14
Cassava	280,838	13,172	4.69	255,020	306,656	4.37

3. Survey Estimate of Total Number of Farm Households, by Province

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
ZAMBIA	897,148	19,389	2.16	859,145	935,150	******	6154
PROVINCE							
Central	83,689	4,625	5.53	74,623	92,754	1.93	651
Copperbelt	38,602	4,773	12.37	29,247	47,958	4.23	332
Eastern	194,121	7,312	3.77	179,790	208,452	2.41	1,191
Luapula	124,278	10,372	8.35	103,949	144,608	6.89	754
Lusaka	16,551	3,100	18.73	10,474	22,627	4.06	189
Northern	166,250	8,239	4.96	150,103	182,398	3.44	1,178
Northwestern	56,416	3,411	6.05	49,730	63,103	1.51	420
Southern	113,717	6,620	5.82	100,742	126,692	3.03	810
Western	103,523	6,146	5.94	91,477	115,569	2.83	629

4. Survey Estimate of Total Farm Area in Has., by Province

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
ZAMBIA	1,187,811	35,008	2.95	1,119,196	1,256,426	5.63	6154
PROVINCE							
Central	131,793	12,805	9.72	106,695	156,890	2.16	651
Copperbelt	41,056	4,821	11.74	31,606	50,505	1.94	332
Eastern	293,313	18,999	6.48	256,074	330,551	4.08	1,191
Luapula	125,061	10,181	8.14	105,107	145,015	3.69	754
Lusaka	19,003	7,828	41.19	3,661	34,345	4.39	189
Northern	227,638	11,702	5.14	204,703	250,574	2.02	1,178
Northwestern	55,267	3,925	7.10	47,574	62,960	1.44	420
Southern	184,800	16,080	8.70	153,284	216,316	2.96	810
Western	109,880	10,077	9.17	90,129	129,631	3.56	629

5. Survey Estimate of Total Number of Cattle, by Province

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
ZAMBIA	1,016,499	116,558	11.47	788,046	1,244,952	1.65	6154
PROVINCE							
Central	88,719	25,554	28.80	38,633	138,804	1.44	651
Copperbelt	1,898	1,029	54.20	-118	3,914	0.22	332
Eastern	209,394	24,108	11.51	162,142	256,646	1.56	1,191
Luapula	1,836	1,256	68.41	-626	4,297	1.16	754
Lusaka	26,248	21,079	80.31	-15,067	67,563	1.46	189
Northern	62,623	18,698	29.86	25,974	99,272	1.88	1,178
Northwestern	7,379	3,061	41.48	1,380	13,378	0.65	420
Southern	368,592	94,482	25.63	183,408	553,776	1.65	810
Western	249,811	51,173	20.48	149,511	350,111	1.57	629

6. Survey Estimate of Total Number of Chickens by Province

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
ZAMBIA	4,307,260	158,550	3.68	3,996,501	4,618,019	4.53	6154
PROVINCE							
Central	565,491	55,887	9.88	455,952	675,030	2.15	651
Copperbelt	122,353	12,431	10.16	97,989	146,718	0.81	332
Eastern	805,663	66,657	8.27	675,014	936,311	3.37	1,191
Luapula	570,855	66,440	11.64	440,633	701,077	4.87	754
Lusaka	93,485	38,981	41.70	17,081	169,888	4.67	189
Northern	900,414	65,966	7.33	771,119	1,029,708	2.82	1,178
Northwestern	254,792	29,351	11.52	197,263	312,321	1.69	420
Southern	672,408	65,660	9.76	543,713	801,102	3.16	810
Western	321,800	44,260	13.75	235,050	408,551	3.64	629

7. Survey Estimate of Average Yield (Kg./Ha.) by Crop
Number of observations:6154

Variable	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect
Maize	1,196.67	39.06	3.26	1,120.12	1,273.23	1.86
Sorghum	655.16	32.56	4.97	591.33	718.98	1.59
Rice	784.69	138.47	17.65	513.28	1,056.10	4.05
Millet	696.76	34.68	4.98	628.80	764.73	4.19
Sunflower	498.34	54.22	10.88	392.08	604.61	0.71
Groundnut	362.46	16.03	4.42	331.04	393.88	1.93
Soyabeans	483.79	191.86	39.66	107.75	859.84	5.94
Cotton Seed	894.43	50.30	5.62	795.85	993.01	1.31
Irish Potatoes	1,602.80	437.91	27.32	744.50	2,461.10	1.70
Virginia Tobacco	569.29	106.69	18.74	360.18	778.41	2.41
Burley Tobacco	1,008.36	105.84	10.50	800.91	1,215.81	1.25
Mixed Beans	412.84	23.35	5.65	367.08	458.59	2.33
Ground Beans	440.17	50.45	11.46	341.29	539.05	1.16
Cowpeas	214.60	51.87	24.17	112.93	316.28	1.86
Sweet Potatoes	1,729.11	192.68	11.14	1,351.45	2,106.76	1.05
Cassava	820.25	48.06	5.86	726.05	914.46	4.98

8. Survey Estimate of Average Farm Size in Has., by Province

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
ZAMBIA	1.35	0.03	2.31	1.29	1.41	3.55	6154
PROVINCE							
Central	1.64	0.14	8.46	1.37	1.91	2.64	651
Copperbelt	1.06	0.10	9.15	0.87	1.25	2.39	332
Eastern	1.52	0.06	4.23	1.40	1.65	4.14	1,191
Luapula	1.04	0.06	6.13	0.91	1.16	5.06	754
Lusaka	1.18	0.33	27.76	0.54	1.82	2.59	189
Northern	1.41	0.07	4.72	1.28	1.54	3.91	1,178
Northwestern	0.99	0.04	3.79	0.92	1.06	1.35	420
Southern	1.68	0.11	6.48	1.47	1.89	3.01	810
Western	1.07	0.07	6.93	0.93	1.22	4.35	629

9. Survey Estimate of Average Number of Cattle per Farm Raising Cattle, by Province

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
ZAMBIA	8.77	0.83	9.49	7.14	10.40	1.31	6154
PROVINCE							
Central	7.31	1.57	21.43	4.24	10.38	1.00	651
Copperbelt	2.41	1.79	74.47	-1.11	5.92	0.48	332
Eastern	6.11	0.41	6.71	5.31	6.92	1.03	1,191
Luapula	3.02	0.88	29.07	1.30	4.73	0.52	754
Lusaka	18.27	4.80	26.28	8.86	27.68	0.20	189
Northern	7.00	1.06	15.17	4.92	9.08	0.74	1,178
Northwestern	6.69	2.10	31.38	2.58	10.81	0.75	420
Southern	9.52	2.11	22.11	5.40	13.65	1.35	810
Western	13.90	2.11	15.20	9.76	18.04	1.23	629

10. Survey Estimate of Average Number of Chickens per Farm Raising Chickens, by Province

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
ZAMBIA	7.26	0.16	2.21	6.95	7.58	2.27	6154
PROVINCE							
Central	9.62	0.55	5.71	8.54	10.70	1.46	651
Copperbelt	6.15	0.33	5.30	5.51	6.78	0.50	332
Eastern	6.65	0.38	5.75	5.90	7.40	3.35	1,191
Luapula	6.45	0.33	5.15	5.80	7.10	2.05	754
Lusaka	10.19	1.96	19.20	6.35	14.02	1.72	189
Northern	7.47	0.37	4.90	6.75	8.19	2.80	1,178
Northwestern	8.17	0.41	5.02	7.36	8.97	0.75	420
Southern	7.47	0.45	5.97	6.60	8.35	2.28	810
Western	5.99	0.46	7.75	5.08	6.90	2.26	629

11. Survey Estimates of Total Crop Production in Metric Tons by Category, for Zambia

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
CROP BY CATEGORY							
Maize							
Category A	557,908	24,297	4.36	510,286	605,530	2.58	5,559
Category B	65,672	6,421	9.78	53,087	78,257	0.28	595
Sorghum							
Category A	22,927	2,741	11.96	17,554	28,299	3.81	5,559
Category B	671	161	23.94	356	986	0.31	595
Rice							
Category A	7,463	1,546	20.71	4,433	10,492	2.86	5,559
Category B	73	30	41.22	14	132	0.18	595
Millet							
Category A	43,133	2,972	6.89	37,307	48,958	3.35	5,559
Category B	1,355	260	19.19	845	1,865	0.39	595
Sunflowers							
Category A	5,139	1,353	26.33	2,487	7,792	2.48	5,559
Category B	996	325	32.66	358	1,633	0.19	595
Groundnuts							
Category A	41,151	2,804	6.81	35,654	46,648	3.16	5,559
Category B	2,528	280	11.09	1,979	3,078	0.29	595
Soyabeans							
Category A	2,149	1,117	52.01	-42	4,339	5.24	5,559
Category B	223	104	46.79	18	427	0.33	595
Cotton seed							
Category A	61,916	11,008	17.78	40,340	83,493	7.50	5,559
Category B	9,017	1,922	21.32	5,249	12,785	0.40	595
Irish potatoes							
Category A	2,140	885	41.36	405	3,874	2.46	5,559
Category B	26	19	74.40	-12	64	0.14	595
Virginia tobacco							
Category A	1,095	515	47.08	84	2,105	3.00	5,559
Category B	79	47	59.22	-13	170	0.43	595
Burley tobacco							
Category A	3,228	1,462	45.28	363	6,093	4.29	5,559
Category B	265	215	81.11	-156	687	0.29	595
Mixed beans							
Category A	9,522	1,279	13.43	7,016	12,029	3.00	5,559
Category B	389	81	20.79	231	548	0.20	595

Ground beans							
Category A	1,081	205	19.01	678	1,484	1.68	5,559
Category B	42	15	35.44	13	71	0.18	595
Cowpeas							
Category A	390	132	34.00	130	649	1.55	5,559
Category B	97	62	64.52	-26	219	1.57	595
Sweet potatoes							
Category A	36,959	3,740	10.12	29,628	44,289	2.72	5,559
Category B	1,455	339	23.31	790	2,119	0.33	595
Cassava							
Category A	226,175	15,103	6.68	196,572	255,777	4.43	5,559
Category B	4,184	973	23.25	2,277	6,091	0.47	595

12. Survey Estimates of Total Crop Production in Metric Tons by Farm Size Group, for Zambia

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
CROP BY FARM SIZE							
Maize							
0.01-0.99 ha.	127,622	6,870	5.38	114,157	141,087	2.75	2,639
1.00-1.99 has.	183,966	9,919	5.39	164,524	203,407	2.60	1,894
2.00-2.99 has.	107,638	10,239	9.51	87,570	127,706	1.37	695
3.00-3.99 has.	68,007	8,791	12.93	50,776	85,238	1.65	297
4.00-4.99 has.	43,485	6,542	15.04	30,663	56,306	1.05	160
5.00-9.99 has.	69,167	8,771	12.68	51,976	86,359	0.89	303
10+ has.	23,696	7,759	32.74	8,489	38,903	0.55	46
Sorghum							
0.01-0.99 ha.	5,934	1,043	17.57	3,890	7,978	3.62	2,639
1.00-1.99 has.	9,376	1,354	14.44	6,723	12,030	1.91	1,894
2.00-2.99 has.	5,951	1,181	19.84	3,636	8,265	2.11	695
3.00-3.99 has.	991	236	23.81	529	1,453	0.93	297
4.00-4.99 has.	594	193	32.41	217	972	0.58	160
5.00-9.99 has.	722	256	35.42	221	1,223	0.92	303
10+ has.	29	24	82.93	-18	76	0.30	46
Rice							
0.01-0.99 ha.	1,897	525	27.68	868	2,927	1.90	2,639
1.00-1.99 has.	2,386	575	24.10	1,259	3,512	1.61	1,894
2.00-2.99 has.	1,961	675	34.42	638	3,284	1.87	695
3.00-3.99 has.	1,052	544	51.73	-15	2,118	1.34	297
4.00-4.99 has.	8	8	100.00	-7	23	0.44	160
5.00-9.99 has.	226	207	91.56	-180	632	1.39	303
10+ has.	6	6	100.00	-6	18	0.20	46
Millet							
0.01-0.99 ha.	11,405	1,249	10.95	8,958	13,853	3.30	2,639
1.00-1.99 has.	17,136	1,504	8.78	14,188	20,083	2.51	1,894
2.00-2.99 has.	7,489	1,035	13.82	5,460	9,518	2.09	695
3.00-3.99 has.	4,328	765	17.68	2,828	5,828	1.38	297
4.00-4.99 has.	2,424	668	27.54	1,116	3,733	1.07	160
5.00-9.99 has.	1,624	613	37.73	423	2,825	1.29	303
10+ has.	81	45	55.85	-8	169	0.12	46

12. Survey Estimates of Total Crop Production in Metric Tons by Farm Size Group, for Zambia (Continued)

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
CROP BY FARM SIZE (Continued)							
Sunflowers							
0.01-0.99 ha.	30	15	51.28	-0	60	0.89	2,639
1.00-1.99 has.	914	378	41.36	173	1,654	2.52	1,894
2.00-2.99 has.	1,003	512	51.04	-0	2,006	2.38	695
3.00-3.99 has.	941	399	42.40	159	1,723	1.25	297
4.00-4.99 has.	888	505	56.88	-102	1,878	3.09	160
5.00-9.99 has.	1,789	705	39.39	408	3,170	1.19	303
10+ has.	571	326	57.17	-69	1,211	0.21	46
Groundnuts							
0.01-0.99 ha.	7,281	1,343	18.45	4,648	9,913	2.15	2,639
1.00-1.99 has.	14,106	902	6.40	12,338	15,875	2.23	1,894
2.00-2.99 has.	9,866	1,182	11.98	7,549	12,183	1.81	695
3.00-3.99 has.	4,785	712	14.88	3,390	6,181	2.09	297
4.00-4.99 has.	3,176	707	22.26	1,791	4,562	1.96	160
5.00-9.99 has.	3,673	595	16.19	2,507	4,839	1.04	303
10+ has.	792	279	35.26	244	1,339	0.52	46
Soyabeans							
0.01-0.99 ha.	122	67	54.60	-9	252	1.23	2,639
1.00-1.99 has.	790	436	55.15	-64	1,644	2.80	1,894
2.00-2.99 has.	408	167	40.89	81	735	1.29	695
3.00-3.99 has.	183	141	76.83	-93	459	0.82	297
4.00-4.99 has.	432	388	89.92	-329	1,193	1.45	160
5.00-9.99 has.	409	218	53.24	-18	836	1.06	303
10+ has.	28	28	100.00	-27	82	0.14	46
Cotton seed							
0.01-0.99 ha.	3,710	569	15.33	2,595	4,825	1.05	2,639
1.00-1.99 has.	18,466	2,529	13.69	13,510	23,422	2.67	1,894
2.00-2.99 has.	16,659	2,873	17.25	11,028	22,291	2.50	695
3.00-3.99 has.	9,928	2,279	22.95	5,462	14,395	1.75	297
4.00-4.99 has.	7,071	3,259	46.09	683	13,460	5.24	160
5.00-9.99 has.	12,878	3,640	28.27	5,743	20,012	0.95	303
10+ has.	2,220	694	31.25	860	3,579	0.52	46

12. Survey Estimates of Total Crop Production in Metric Tons by Farm Size Group, for Zambia (Continued)

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
CROP BY FARM SIZE (Continued)							
Irish potatoes							
0.01-0.99 ha.	327	133	40.61	67	587	1.16	2,639
1.00-1.99 has.	1,211	594	49.07	46	2,377	1.61	1,894
2.00-2.99 has.	224	162	72.27	-93	542	1.10	695
3.00-3.99 has.	0	0	******	0	0	******	297
4.00-4.99 has.	22	18	80.38	-13	58	0.13	160
5.00-9.99 has.	321	321	100.00	-308	949	1.83	303
10+ has.	60	60	100.00	-58	178	0.92	46
Virginia tobacco							
0.01-0.99 ha.	69	69	100.00	-67	205	1.58	2,639
1.00-1.99 has.	569	336	58.97	-89	1,227	2.52	1,894
2.00-2.99 has.	58	58	100.00	-56	173	1.45	695
3.00-3.99 has.	145	102	70.63	-56	346	1.82	297
4.00-4.99 has.	10	10	100.00	-10	29	0.14	160
5.00-9.99 has.	322	209	64.94	-88	731	1.17	303
10+ has.	0	0	******	0	0	******	46
Burley tobacco							
0.01-0.99 ha.	40	34	84.65	-27	107	1.03	2,639
1.00-1.99 has.	809	377	46.62	70	1,548	1.56	1,894
2.00-2.99 has.	951	705	74.14	-431	2,333	3.01	695
3.00-3.99 has.	685	383	55.90	-66	1,435	1.18	297
4.00-4.99 has.	576	422	73.24	-251	1,404	1.99	160
5.00-9.99 has.	432	432	100.00	-414	1,278	1.01	303
10+ has.	0	0	******	0	0	******	46
Mixed beans							
0.01-0.99 ha.	1,231	198	16.12	842	1,620	1.38	2,639
1.00-1.99 has.	3,151	436	13.85	2,296	4,007	1.96	1,894
2.00-2.99 has.	2,429	369	15.19	1,706	3,152	1.37	695
3.00-3.99 has.	1,444	338	23.41	781	2,106	1.60	297
4.00-4.99 has.	733	220	30.05	301	1,165	0.93	160
5.00-9.99 has.	385	121	31.50	147	622	0.77	303
10+ has.	539	473	87.73	-388	1,466	1.01	46

12. Survey Estimates of Total Crop Production in Metric Tons by Farm Size Group, for Zambia (Continued)

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
CROP BY FARM SIZE (Continued)							
Ground beans							
0.01-0.99 ha.	113	35	31.28	44	182	1.10	2,639
1.00-1.99 has.	410	99	24.02	217	603	1.61	1,894
2.00-2.99 has.	176	65	37.01	48	304	1.27	695
3.00-3.99 has.	203	94	46.23	19	387	1.65	297
4.00-4.99 has.	138	72	52.23	-3	279	0.90	160
5.00-9.99 has.	82	58	70.12	-31	195	0.69	303
10+ has.	0	0	100.00	-0	1	0.06	46
Cowpeas							
0.01-0.99 ha.	66	42	64.02	-17	150	2.87	2,639
1.00-1.99 has.	99	59	60.06	-18	215	3.06	1,894
2.00-2.99 has.	59	52	88.02	-43	160	1.86	695
3.00-3.99 has.	95	61	64.24	-25	214	1.08	297
4.00-4.99 has.	78	50	64.23	-20	176	1.28	160
5.00-9.99 has.	90	62	69.15	-32	211	0.74	303
10+ has.	0	0	******	0	0	******	46
Sweet potatoes							
0.01-0.99 ha.	8,130	1,122	13.81	5,930	10,330	2.14	2,639
1.00-1.99 has.	14,547	1,890	12.99	10,844	18,251	2.22	1,894
2.00-2.99 has.	8,527	1,561	18.31	5,467	11,588	1.36	695
3.00-3.99 has.	3,457	898	25.98	1,697	5,218	1.26	297
4.00-4.99 has.	1,187	333	28.06	534	1,840	0.84	160
5.00-9.99 has.	2,290	808	35.26	708	3,873	0.85	303
10+ has.	273	146	53.54	-14	560	0.18	46
Cassava							
0.01-0.99 ha.	72,252	6,546	9.06	59,422	85,083	3.51	2,639
1.00-1.99 has.	89,161	7,558	8.48	74,348	103,974	3.44	1,894
2.00-2.99 has.	42,203	7,259	17.20	27,975	56,432	2.38	695
3.00-3.99 has.	16,088	2,979	18.52	10,249	21,927	2.21	297
4.00-4.99 has.	5,347	1,291	24.15	2,815	7,878	1.19	160
5.00-9.99 has.	5,024	1,966	39.13	1,171	8,878	1.35	303
10+ has.	283	179	63.19	-67	633	0.28	46

13. Survey Estimates of Average Crop Yield per Household Growing Crop (Kgs./Ha.), for Zambia

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
CROP							
Maize	1,238.78	31.54	2.55	1,176.96	1,300.61	3.26	4,348
Sorghum	762.24	40.23	5.28	683.39	841.09	2.62	587
Rice	961.12	108.66	11.31	748.15	1,174.10	2.55	157
Millet	828.80	27.83	3.36	774.24	883.35	2.93	1,111
Sunflowers	538.01	48.67	9.05	442.61	633.40	0.77	164
Groundnuts	432.32	31.00	7.17	371.55	493.09	1.92	2,428
Soyabeans	541.54	172.98	31.94	202.49	880.58	4.70	82
Cotton seed	945.61	42.85	4.53	861.63	1,029.60	2.29	625
Irish potatoes	1,995.68	481.28	24.12	1,052.38	2,938.99	1.50	31
Virginia tobacco	712.28	185.16	26.00	349.37	1,075.19	2.45	21
Burley tobacco	1,008.17	95.28	9.45	821.42	1,194.93	0.98	29
Mixed beans	424.84	23.22	5.47	379.33	470.35	2.63	569
Ground beans	512.82	71.88	14.02	371.94	653.69	1.63	105
Cowpeas	210.94	44.88	21.28	122.97	298.91	2.25	44
Sweet potatoes	2,296.52	144.09	6.27	2,014.09	2,578.94	3.34	661
Cassava	1,251.74	67.22	5.37	1,120.00	1,383.49	4.28	1,714

14. Survey Estimates of Average Crop Yield per Household Growing Crop (Kgs./Ha.), by Farm Size

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
CROP BY FARM SIZE							
Maize							
0.01-0.99 has.	1,261.80	45.68	3.62	1,172.28	1,351.32	2.37	1,716
1.00-1.99 has.	1,196.55	38.32	3.20	1,121.44	1,271.66	2.45	1,357
2.00-2.99 has.	1,157.73	63.92	5.52	1,032.44	1,283.01	1.73	548
3.00-3.99 has.	1,334.25	106.23	7.96	1,126.05	1,542.45	1.81	252
4.00-4.99 has.	1,462.35	165.12	11.29	1,138.71	1,785.99	1.73	140
5.00-9.99 has.	1,387.60	125.12	9.02	1,142.37	1,632.84	1.15	289
10+ has.	1,552.74	435.57	28.05	699.03	2,406.46	0.83	46
Sorghum							
0.01-0.99 has.	740.68	59.74	8.07	623.58	857.77	2.35	199
1.00-1.99 has.	797.25	59.67	7.48	680.30	914.20	1.82	215
2.00-2.99 has.	782.95	76.36	9.75	633.28	932.62	1.98	106
3.00-3.99 has.	674.27	102.37	15.18	473.62	874.92	0.87	23
4.00-4.99 has.	843.54	178.51	21.16	493.65	1,193.43	1.10	15
5.00-9.99 has.	522.48	149.79	28.67	228.89	816.06	0.62	27
10+ has.	406.44	49.68	12.22	309.06	503.81	0.14	2
Rice							
0.01-0.99 has.	983.08	189.76	19.30	611.15	1,355.00	3.01	58
1.00-1.99 has.	959.29	172.11	17.94	621.96	1,296.62	2.77	54
2.00-2.99 has.	998.53	219.32	21.96	568.67	1,428.39	1.32	29
3.00-3.99 has.	894.89	310.07	34.65	287.16	1,502.62	1.31	12
4.00-4.99 has.	120.00	0.00	0.00	120.00	120.00	0.00	1
5.00-9.99 has.	324.60	5.75	1.77	313.32	335.87	0.34	2
10+ has.	500.00	0.00	0.00	500.00	500.00	0.00	1
Millet							
0.01-0.99 has.	849.69	36.92	4.35	777.33	922.05	1.64	354
1.00-1.99 has.	845.67	37.31	4.41	772.54	918.79	2.49	447
2.00-2.99 has.	733.45	45.00	6.14	645.24	821.66	1.25	156
3.00-3.99 has.	834.95	71.45	8.56	694.90	975.00	1.07	78
4.00-4.99 has.	823.99	97.31	11.81	633.26	1,014.72	0.65	33
5.00-9.99 has.	772.63	194.51	25.18	391.38	1,153.87	0.86	39
10+ has.	564.54	195.01	34.54	182.33	946.76	0.25	4

14. Survey Estimates of Average Crop Yield per Household Growing Crop (Kgs./Ha.), by Farm Size (Continued)

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
CROP BY FARM SIZE (Continued)							
Sunflowers							
0.01-0.99 has.	217.40	68.30	31.42	83.53	351.28	1.09	6
1.00-1.99 has.	551.12	87.87	15.94	378.89	723.34	1.48	35
2.00-2.99 has.	514.77	72.08	14.00	373.49	656.04	0.89	21
3.00-3.99 has.	699.52	101.01	14.44	501.54	897.51	0.44	20
4.00-4.99 has.	621.96	151.44	24.35	325.14	918.78	2.24	22
5.00-9.99 has.	487.17	110.55	22.69	270.49	703.85	0.26	49
10+ has.	393.06	160.22	40.76	79.02	707.10	0.23	11
Groundnuts							
0.01-0.99 has.	459.82	99.93	21.73	263.96	655.68	1.94	607
1.00-1.99 has.	404.99	14.30	3.53	376.95	433.02	1.83	884
2.00-2.99 has.	442.04	33.96	7.68	375.48	508.61	1.49	392
3.00-3.99 has.	413.41	31.78	7.69	351.11	475.70	1.16	190
4.00-4.99 has.	463.34	47.67	10.29	369.91	556.77	0.89	104
5.00-9.99 has.	488.14	91.48	18.74	308.84	667.45	1.14	214
10+ has.	328.80	74.61	22.69	182.56	475.05	0.70	37
Soyabeans							
0.01-0.99 has.	375.37	126.38	33.67	127.66	623.08	1.43	9
1.00-1.99 has.	543.72	230.16	42.33	92.61	994.84	3.51	28
2.00-2.99 has.	585.38	242.09	41.36	110.90	1,059.87	1.62	19
3.00-3.99 has.	561.31	252.60	45.00	66.21	1,056.41	0.83	7
4.00-4.99 has.	832.52	415.98	49.97	17.19	1,647.85	1.27	7
5.00-9.99 has.	415.62	120.58	29.01	179.28	651.96	1.02	11
10+ has.	337.50	0.00	0.00	337.50	337.50	******	1
Cotton seed							
0.01-0.99 has.	1,116.26	91.25	8.17	937.42	1,295.11	1.13	62
1.00-1.99 has.	966.80	65.34	6.76	838.72	1,094.87	2.33	184
2.00-2.99 has.	870.00	62.23	7.15	748.03	991.97	1.44	139
3.00-3.99 has.	983.09	111.70	11.36	764.15	1,202.03	1.47	62
4.00-4.99 has.	882.10	106.86	12.11	672.66	1,091.55	1.95	46
5.00-9.99 has.	889.73	132.15	14.85	630.73	1,148.74	0.59	109
10+ has.	432.58	75.16	17.37	285.27	579.88	0.39	23

14. Survey Estimates of Average Crop Yield per Household Growing Crop (Kgs./Ha.), by Farm Size (Continued)

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
CROP BY FARM SIZE (Continued)							
Irish potatoes							
0.01-0.99 has.	1,860.13	654.16	35.17	577.99	3,142.28	1.33	8
1.00-1.99 has.	2,077.89	645.99	31.09	811.75	3,344.03	0.95	15
2.00-2.99 has.	897.49	406.84	45.33	100.09	1,694.89	1.26	4
4.00-4.99 has.	1,678.81	502.35	29.92	694.21	2,663.41	0.11	2
5.00-9.99 has.	3,243.24	0.00	0.00	3,243.23	3,243.25	0.94	1
10+ has.	4,500.00	0.01	0.00	4,499.99	4,500.01	******	1
Virginia tobacco							
0.01-0.99 has.	493.44	0.00	0.00	493.44	493.44	0.31	1
1.00-1.99 has.	655.89	235.41	35.89	194.48	1,117.30	1.41	8
2.00-2.99 has.	1,374.28	0.00	0.00	1,374.28	1,374.28	0.00	2
3.00-3.99 has.	886.65	253.71	28.61	389.39	1,383.92	1.59	4
4.00-4.99 has.	1,250.00	0.00	0.00	1,250.00	1,250.00	******	1
5.00-9.99 has.	530.74	93.68	17.65	347.13	714.35	0.72	5
Burley tobacco							
0.01-0.99 has.	829.39	182.66	22.02	471.37	1,187.40	1.15	2
1.00-1.99 has.	1,056.76	173.07	16.38	717.54	1,395.97	0.75	9
2.00-2.99 has.	785.15	109.51	13.95	570.52	999.78	0.56	8
3.00-3.99 has.	1,190.24	226.68	19.04	745.95	1,634.53	1.24	4
4.00-4.99 has.	1,163.50	74.26	6.38	1,017.94	1,309.05	1.61	2
5.00-9.99 has.	1,149.01	0.00	0.00	1,149.01	1,149.01	0.00	4
Mixed beans							
0.01-0.99 has.	419.86	36.59	8.71	348.14	491.57	1.33	136
1.00-1.99 has.	394.45	30.38	7.70	334.90	454.00	2.07	227
2.00-2.99 has.	435.92	35.34	8.11	366.65	505.19	1.44	96
3.00-3.99 has.	508.92	54.49	10.71	402.13	615.72	0.97	51
4.00-4.99 has.	532.01	84.44	15.87	366.52	697.51	1.31	22
5.00-9.99 has.	462.83	62.60	13.52	340.14	585.51	0.62	32
10+ has.	421.37	72.25	17.15	279.77	562.97	0.24	5

14. Survey Estimates of Average Crop Yield per Household Growing Crop (Kgs./Ha.), by Farm Size (Continued)

Category	Estimate	Standard Error	C.V. (%)	95% Confidence Interval Lower	Upper	Design Effect	Number of Observations
CROP BY FARM SIZE (Continued)							
Ground beans							
0.01-0.99 has.	331.74	126.10	38.01	84.58	578.90	2.60	20
1.00-1.99 has.	473.39	83.51	17.64	309.72	637.06	1.28	44
2.00-2.99 has.	582.17	99.79	17.14	386.58	777.76	0.62	16
3.00-3.99 has.	696.60	145.98	20.96	410.49	982.71	1.33	11
4.00-4.99 has.	525.55	80.71	15.36	367.36	683.75	0.63	6
5.00-9.99 has.	1,814.81	924.70	50.95	2.40	3,627.23	0.60	7
10+ has.	750.00	0.00	0.00	750.00	750.00	0.00	1
Cowpeas							
0.01-0.99 has.	152.09	61.53	40.46	31.48	272.69	2.62	6
1.00-1.99 has.	129.25	49.15	38.03	32.91	225.59	2.18	11
2.00-2.99 has.	594.14	99.06	16.67	399.98	788.30	0.52	5
3.00-3.99 has.	338.41	68.45	20.23	204.24	472.57	1.15	5
4.00-4.99 has.	227.52	89.96	39.54	51.19	403.85	0.89	8
5.00-9.99 has.	365.63	114.95	31.44	140.34	590.92	0.53	9
Sweet potatoes							
0.01-0.99 has.	2,205.27	183.90	8.34	1,844.83	2,565.72	2.00	204
1.00-1.99 has.	2,258.97	178.01	7.88	1,910.07	2,607.88	2.03	243
2.00-2.99 has.	2,473.08	301.84	12.21	1,881.47	3,064.69	2.06	106
3.00-3.99 has.	2,218.46	287.61	12.96	1,654.76	2,782.17	0.81	40
4.00-4.99 has.	2,983.69	464.70	15.57	2,072.89	3,894.49	1.03	22
5.00-9.99 has.	2,313.05	498.90	21.57	1,335.21	3,290.90	0.75	40
10+ has.	3,460.66	468.91	13.55	2,541.60	4,379.73	0.24	6
Cassava							
0.01-0.99 has.	1,638.86	114.52	6.99	1,414.40	1,863.33	3.16	641
1.00-1.99 has.	1,075.53	58.51	5.44	960.85	1,190.20	2.84	654
2.00-2.99 has.	911.00	112.36	12.33	690.77	1,131.22	1.92	239
3.00-3.99 has.	721.67	74.63	10.34	575.39	867.95	1.20	98
4.00-4.99 has.	621.20	106.51	17.15	412.44	829.96	1.01	36
5.00-9.99 has.	832.39	327.48	39.34	190.53	1,474.25	1.90	42
10+ has.	911.66	325.50	35.70	273.69	1,549.64	0.24	4