Mission Report

for a short-term mission of the specialist in sampling for household surveys

From 21 March to 11 April **2015**

within the framework of the

AGREEMENT ON CONSULTING ON
INSTITUTIONAL CAPACITY BUILDING,
ECONOMIC STATISTICS AND RELATED AREAS

between
INE and Scanstat

David J. Megill

INSTITUTO NACIONAL DE ESTATÍSTICA

.

Address in U.S.A.:   *David J. Megill*
                          1504 Kenwood Ave.
                          Alexandria, VA 22302
E-Mail: *davidmegill@yahoo.com*
Telephone: *1-703-824-0292*

Address in Mozambique   *Hotel Terminus*
                          Maputo
Telephone: *+258-82-963-9620*

Table of Contents

1. INTRODUCTION AND TERMS OF REFERENCE

The *Instituto Nacional de Estatística* (INE) is conducting the *Inquérito sobre o Orçamento Familiar* (IOF) 2014/15, or Household Budget Survey (HBS), in a nationally-representative sample of 11,592 households in 1,236 sample census enumeration areas (EAs) over the 12-month period from August 2014 to July 2015. This survey was designed as a combination of the *Inquérito Contínuo de Agregados Familiares* (INCAF), or Continuous Household Survey, which is a multipurpose household survey with a quarterly employment component, and the HBS, designed to obtain income and expenditure data for all four quarters to represent seasonality. One of the objectives of the IOF is to provide measures of poverty and other socioeconomic indicators, and to provide information on consumption needed for national accounts. The cross-sectional survey data from the full sample of households each quarter can be used to provide current estimates of key indicators such as the unemployment rate. In addition, the sample of households for IOF can be treated as a panel, since each sample household is interviewed each quarter in a different period of the month; this will ensure that the survey data are representative of the household income and expenditures over time.

The first quarter of data collection for IOF was conducted between 8 August and 7 November 2014, and the second quarter was completed on 7 February. Following the data collection and processing for the first quarter of IOF, the Scanstat short-term consultants assisted INE in calculating the weights for the IOF data from the first quarter, reviewing data quality and producing preliminary results. The purpose of this second mission in April 2015 is to follow up with the calculation of weights and the analysis of the first two quarters of IOF data. David Megill, the Scanstat Sampling Consultant, began his mission on 23 March 2015, in order to review the data for the second quarter and calculate the corresponding weights. The other Scanstat consultants, Lars Lundgren and Anne Abelseth, arrived two weeks later to provide technical assistance and training for the analysis and data processing for the IOF second quarter.

The original Terms of Reference for Megill's second mission were stated as follows:

- **Objective:** During the second mission the Sampling Consultant will follow up the findings and recommendations from the first visit, and review the panel data from the first two quarters of INCAF/IOF.
- **Activities:** The weights for the second quarter of INCAF/IOF data will be finalized, and the weighting procedures for using combined data for different quarters will be developed. Sampling errors and design effects will be tabulated for key survey estimates from the second quarter of INCAF/IOF data, as well as for the combined data for the first two quarters of INCAF/IOF. The effect of the panel methodology on improving the estimates of quarterly differences in the unemployment rate and other key indicators will be studied. The effect of the panel methodology on the cross-sectional estimates of quarterly household income and expenditure data will also be examined.
- **Expected outputs:** The Sampling Consultant and INE counterparts will conduct a workshop on the IOF sample design and estimation procedures, including the methodology for the panel survey, sample rotation and partial overlap of the sample each quarter. The maintenance of the sampling frame over time will also be reviewed.
- **Reporting:** The Sampling Consultant will submit a second report on findings and recommendations for the INCAF/IOF sampling and estimation methodology.

The actual scope of work for this mission was modified to address the priority of assisting INE with corrections to the second quarter data file and calculating the final weights, as described in this report. The main activity of this mission, the calculation of the IOF weights, was only completed remotely following this mission once the clean IOF data file for the second quarter was available. The calculation of sampling errors for selected IOF indicators can only be accomplished once the IOF data edits are complete and the weighted survey data file for the second quarter is considered final. Therefore this activity will be followed up during the third mission of the Sampling Consultant.

This report includes detailed documentation of the weighting procedures for the cross-sectional IOF data from the second quarter and for the first six months, as well as the weighting procedures for the panel data from the first and second quarters. These weighting procedures depend on the IOF sample design, which is documented in Megill's December 2014 Mission Report.

During this mission Megill worked closely with Arão Balate, Director, *Direcção de Censos e Inquéritos*, Basílio Cubula, INE Sampling Statistician, and other INE staff in implementing the weighting procedures for the IOF 2014/15. He also collaborated with his Scanstat consultant colleagues, Lars Carlsson, Lars Lundgren and Anne Abelseth. He appreciates their collaboration, and he would also like to thank Dr. João Loureiro, INE President, Manuel Gaspar, INE Vice-President, Antônio Adriano, Director Adjunto, *Direcção de Censos e Inquéritos*, and Cristóvão Muahio, Chief, *Departamento de Metodologia e Amostragem* (DMA), for their support.

2. ACTIVITIES DURING THE MISSION

At the beginning of this mission Megill met with Arão Balate, Basílio Cubula and other INE staff to discuss the status of the IOF data collection and processing. Unfortunately the IOF data collection had stopped soon after the second quarter given that the budget could not be released because of political issues in the Government of Mozambique. This will result in a serious seasonal gap in the IOF data for the third quarter, especially since this period generally has a higher level of agricultural production. Therefore it will be necessary to discuss how to handle this gap in planning for the remainder of the IOF data collection and the analysis.

Of the original 1,236 sample EAs selected for IOF, 1,233 were covered during the first quarter, and 1,175 of these EAs were enumerated during the second quarter. It was not possible to enumerate 58 of the 1,233 enumeration areas covered in the first quarter; the main problem was in the province of Zambézia, where 50 EAs (9 urban and 41 rural) could not be enumerated in the second quarter because of a major flood.

2.1. Review and Correction of the IOF Data for the Second Quarter

The original IOF household data for each quarter were collected in the field using a CSPro CAPI (computer-assisted personal interviewing) application on tablet computers. The data files for the individual clusters were sent to INE and concatenated into a complete IOF data file for the quarter. The full CSPro data file was then used to export SPSS files with the IOF household and employment data. The income and expenditure data were captured in a separate data file; these data were originally collected using a

paper questionnaire, which was then entered on a tablet in the field. Finally the income and expenditure data from the paper questionnaires were entered again in the central office in order to verify the data entry from the field. Unfortunately this process results in a corresponding delay in the availability of the income and expenditure data from IOF, so this operation needs to be streamlined. At the time of this mission only 50% of the IOF income and expenditure for the second quarter had been verified.

The INE data processing staff exported an SPSS version of the IOF household data file from the full concatenated CSPro data file, with a record for each person in the sample households. They provided Megill with this SPSS data file for the second quarter during the first week of his visit. Megill used this SPSS data file to produce aggregated data at the sample household and cluster levels to verify the concatenation of the data and determine which sample clusters were not enumerated. First it was found that there were 6 sample EAs identified as urban in the IOF data file that were actually rural according to the sampling frame (in the provinces of Gaza and Maputo). The corresponding data were corrected. However, it was also found that the same urban/rural classification errors appeared in the final version of the IOF data for the first quarter. This resulted in a slight bias in the weights and preliminary urban/rural results from the first quarter. Therefore Megill worked with the INE staff in revising the household weights for the first quarter. These revised first quarter weights were provided to the other Scanstat consultants.

In reviewing the SPSS version of the the IOF household data, Megill found that the final interview status was blank for more than 18% of the household records. Apparently there was a problem with the CAPI application for entering the IOF data on tablets. At the end of the interview the program was ended without requiring the interviewer to enter the final status of the interview (completed, partially completed, not at home, refused, etc.). The INE data processing staff were aware of this problem and had already begun imputing code 1 or 2 (complete or incomplete, respectively) for the final interview status based on the IOF data that appeared in the corresponding records. Since the final interview status was needed to determine the final number of completed interviews in each sample cluster for calculating the weights, Megill provided a list of the households without a final interview status, and the INE staff completed the imputation of the corresponding values.

Another problem that was found in the IOF data file for the second quarter is that some of the IOF sample household identification numbers (ID07) from the first quarter had been changed, making it impossible to match some of the sample households from the two quarters. Apparently this occured in some clusters when the first household had ID07 equal to 02; in these clusters the original household 01 could not be interviewed in the first quarter. Because of an error in the program, the interviewer had to change the code of the first household to 01 in order for the tablet data entry application to work, and this changed the codes for the other households in the cluster. Megill provided the INE staff with a list of the sample households that appeared in the data for the second quarter, but not in the first quarter, which were mostly related to this problem. In this case the INE staff made the corresponding corrections. The INE data processing staff had also discovered that some of the sample households that were included in the CSPro data file from the field did not appear in the file exported to SPSS. They starting adding the missing households to the SPSS data file, which was time-consuming.

When Anne Abelseth, Scanstat IT Consultant, arrived on 7 April, she was able to quickly find the source of the problems with the tablet data entry application that were causing the various errors, and she made corrections to the application for the next quarter of IOF. She also discovered that the CSPro data dictionary that the INE staff had used for exporting the SPSS data files had a problem that resulted in some sample households not being exported, and she also corrected this problem. Then Abelseth assisted with the correction of the household identification numbers and the final interview status in the original CSPro file, and she standardized the CSPro data dictionary to avoid such problems in the future.

Since Megill spent much of his time during this mission resolving the IOF data problems, and the final IOF data were not available before his departure, he agreed to work on the final calculation of the IOF second quarter weights and the panel weights remotely the following week. Once he received from Abelseth a summary household-level IOF data file on 17 April, he gave priority to this work and submitted the final weights for the IOF second quarter data on 19 April, as well as weights for the panel of households with completed interviews for both the first and second quarters.

Table 1 shows the distribution of the enumerated sample EAs and households for the first quarter of IOF by province, urban and rural stratum following the correction of the urban/rural codes, and Table 2 shows the corresponding distribution for the second quarter.

Table 1.   Distribution of Enumerated Sample EAs and Households with Completed

Interviews for the First Quarter of IOF 2014/15, by Province and Urban/Rural Stratum

	Urban		Rural		Total	
Province	No. of EAs	No. of Households	No. of EAs	No. of Households	No. of EAs	No. of Households
Niassa	32	352	63	497	95	849
Cabo Delgado	44	467	60	468	104	935
Nampula	60	637	104	810	164	1,447
Zambézia	52	556	124	977	176	1,533
Tete	40	436	68	539	108	975
Manica	40	436	57	445	97	881
Sofala	60	651	41	327	101	978
Inhambane	40	431	52	416	92	847
Gaza	40	438	48	384	88	822
Maputo Província	60	658	48	383	108	1,041
Maputo Cidade	100	1,058	0	0	100	1,058
Total	568	6,120	665	5,246	1,233	11,366

Table 2.   Distribution of Enumerated Sample EAs and Households with Completed

Interviews for the Second Quarter of IOF 2014/15, by Province and Urban/Rural Stratum

Province	Urban		Rural		Total	
	No. of EAs	No. of Households	No. of EAs	No. of Households	No. of EAs	No. of Households
Niassa	32	321	63	448	95	769
Cabo Delgado	44	422	57	426	101	848
Nampula	60	592	104	764	164	1,356
Zambézia	43	419	83	616	126	1,035
Tete	39	403	65	486	104	889
Manica	40	413	56	435	96	848
Sofala	60	647	41	326	101	973
Inhambane	40	419	52	396	92	815
Gaza	40	425	48	373	88	798
Maputo Província	60	641	48	361	108	1,002
Maputo Cidade	100	973	0	0	100	973
Total	558	5,675	617	4,631	1,175	10,306

2.2. Calculation of Weights for IOF Data

The weighting procedures for the IOF 2014/15 depend on the sample design, so first it is necessary to review the sampling methodology used for the survey. This sample design is described in Megill's Scanstat Mission Report of December 2014. A multi-stage sample design based on the master sampling frame was used for IOF. The sampling frame was stratified by province, urban and rural strata. The sample EAs were selected systematically with probability proportional to size (PPS) within each stratum. A sample of 11 households was selected from the listing for each sample urban EA, and 8 households were selected for each rural EA.

In developing the weighting procedures for each set of IOF data, it is important to understand the nature of the sample for the particular analysis that is being planned. Survey data can generally be classified into two major groups: cross-sectional and panel data. In the case of a cross-sectional survey, the objective is to represent the current household-based population over the period of the data collection. For example, since one objective of the IOF is to produce quarterly estimates of the unemployment rate and other key labor force indicators, these estimates should represent the current household-based population each quarter, so the survey data would be treated as cross-sectional. In this case each quarterly survey is considered a separate cross-sectional sample for analyzing these types of current indicators. Since the IOF data collection is based on following a sample of households enumerated in the first quarter for the following three quarters, the sampling procedures do not follow a strictly cross-sectional design, but we use the data for all households with completed interviews regardless of whether they appear in the other quarters. Therefore the cross-sectional weights for each quarter are based on the households with completed interviews for that quarter.

In the case of sample households from the first quarter that move out but another household moves into the same dwelling unit in one of the following quarters, this new household is enumerated for IOF. However, if the sample dwelling unit is vacant or the household refuses, no replacement household is selected after the first quarter. Any new persons found in the sample households after the first quarter are not enumerated. Therefore the effective number of sample households and persons decreases slightly each quarter. This introduces a bias in the cross-sectional estimates, but it is expected that this bias is small as long as the changes in the sample households are relatively minor.

For a panel survey, the sample households in the first quarter are enumerated each following quarter so that the data from all quarters can be linked for a longitudinal analysis. Since it is necessary to link the data for each sample household from all quarters, only the households that have complete interviews for all quarters are included in the analysis. Therefore it is necessary to calculate weights based on the sample households with data for all quarters, so the panel weights will be different from the cross-sectional weights for each quarter.

The weighting procedures for the first quarter IOF data were described in Megill's Mission Report of December 2014. The first quarter of IOF established the panel households that would be followed, and only cross-sectional weights were calculated for that quarter. In the case of the second quarter of IOF, similar cross-sectional weights were calculated for the full set of data for all the households with completed interviews in the second quarter. In order to produce 6-month cross-sectional estimates such as

the unemployment rate, it will be possible to combine all the data from the first and second quarters; in this case it is necessary to calculate a new set of weights for the combined cross-sectional data.

For the panel analysis, it is necessary to link the sample households from the first and second quarters. Therefore a different set of panel weights is calculated for the households with completed interviews for both the first and second quarters. The procedures for calculating the weights for the IOF cross-sectional and panel data are described separately below.

2.2.1. Calculation of Cross-Sectional Weights for the Second Quarter of IOF

The weighting procedures for the second quarter of IOF data are similar to those used for the first quarter. These weighting procedures are described in Megill's Mission Report of December 2014, which also includes a description of the IOF 2014/15 sample design. That report discusses the problem of missing information on the segmenting of large sample EAs and combining of small sample EAs, which resulted in the need to calculate approximate weights.

The weights depend on the final number of enumerated sample EAs in each stratum, as well as the number of completed household interviews in each sample EA. The weighting formula presented in the December Mission Report automatically adjusts the weights for any nonresponse. Since there was no replacement of non-interview panel households beginning with the second quarter, the number of completed interviews each quarter will generally decrease slightly. The cross-sectional weights for the IOF data each quarter are designed to produce estimates that represent the average for each indicator over the 3-month period.

As mentioned above, it was necessary to calculate approximate weights since some of the information needed to determine the exact probabilities was missing. The final weight for the quarterly cross-sectional data was simplified into the following formula:

$$W''_{hij} = \frac{M_h}{n''_h \times m'_{hij}},$$

where:

W''_{hij} = approximate adjusted basic weight for the sample households in the j-th sample EA of the i-th sample PSU in stratum h for the second quarter of IOF

M_h = total number of households in the 2007 Census frame for stratum h

n''_h = number of EAs enumerated in stratum h for the IOF second quarter

m'_{hij} = number of sample households with completed interviews in the j-th sample EA of the i-th sample PSU in stratum h (for the second quarter)

It can be seen in this formula that the final adjusted weight is similar for all sample households within each stratum, varying only by the number of completed household interviews in each sample EA. For the second quarter of IOF, only the households with completed interviews in the first quarter were enumerated, so the weights within a stratum were slightly more variable than those for the first quarter of IOF.

Since the weights depend on the number of sample EAs enumerated in each stratum and the number of households with completed interviews in each sample EA, the first step involved aggregating the IOF household data file for the second quarter by EA, counting the number of households with completed interviews in each sample EA. For this reason it is necessary for the IOF data file to have the correct final interview status for each household. The EA summary file from the final IOF household data for the second quarter included a total of 1,175 EAs and 10,306 households with completed interviews. The distribution of these sample EAs and completed households for the second quarter is shown in Table 2 above. A copy of the spreadsheet used for calculating the weights from the first quarter was adapted for the second quarter cross-sectional weights, since the information from the frame does not change. However, first it was necessary to identify and separate the 1,175 sample EAs that were enumerated in the second quarter. Then the information on the number of enumerated EAs in each stratum was entered into this weighting spreadsheet, as well as the number of households with completed interviews in each sample EA. The weighting spreadsheet includes formulas that automatically calculated the basic weights.

The next step involved adjusting the basic weights based on the population projections, similar to the procedure that was used for the first quarter IOF weights. As described in Megill's December 2014 Mission Report, the adjusted basic weights for the IOF sample households will provide a weighted distribution by province, urban and rural stratum that is consistent with the 2007 Mozambique Census (*Recenseamento Geral da População e Habitação*, RGPH). In order to reflect the growth in the population by stratum between 2007 and the time of the IOF 2014/15 second quarter data collection, the preliminary weights were adjusted based on population projections. In this case it was necessary to interpolate the population projections to the mid-point of the second quarter, or 23 December 2014.

The weight adjustment factor based on the projected total population by province, urban and rural stratum can be expressed as follows:

$$A_h = \frac{P_h}{\sum_{i \in h} \sum_{j} \sum_{k} W''_{hij} \times p_{hijk}} ,$$

where:

$A_{2h} =$ adjustment factor for the basic weights of the IOF sample households in stratum (province, urban/rural) h for the second quarter

$P_{2h} =$ projected total population for stratum h for the mid-point of the data collection period for the second quarter of IOF, based on demographic analysis

$W''_{hij} =$ adjusted second quarter IOF basic cross-sectional weight for the sample households in the j-th sample EA of the i-th sample PSU in stratum h

p_{hijk} = number of persons in the k-th sample household in the j-th sample EA of the i-th sample PSU in stratum h for the second quarter

The denominator of the adjustment factor A_h is the estimated weighted total population in stratum h from the IOF data for the second quarter using the preliminary basic design weights. The preliminary weights for all the sample households within a stratum were multiplied by the corresponding adjustment factor for the stratum to obtain the final adjusted weights, as follows:

$$W_{A2hij} = W''_{hij} \times A_{2h} ,$$

where:

W_{A2hij} = final adjusted weight for the sample households in the j-th sample EA of the i-th sample PSU in stratum h for the second quarter of IOF

After the adjustment factors were applied to the weights of each stratum, the final weighted survey estimates of total population by stratum were consistent with the corresponding population projections for the second quarter. Of course the accuracy of the estimates of total population based on the adjusted weights depends on the quality of the population projections by stratum.

The population projections which INE generated for each year reflect the mid-point of the year, or 1 July. For the adjustment of the weights, it is ideal to use the population projections for the mid-point of the data collection period for the survey. In the case of the second quarter of IOF, the data collection was conducted between 8 November 2014 and 7 February 2015, so the mid-point was estimated as 23 December 2014. Using the population projections by province, urban and rural stratum for 1 July 2014 and 1 July 2015, an interpolation based on exponential growth was used to estimate the population for 23 December 2014, using the following formula:

$$P_h = P_{14h} \times e^{\ln\left[\left(\frac{P_{15h}}{P_{14h}}\right)\times\left(\frac{t_{IOF}-t_{14}}{t_{15}-t_{14}}\right)\right]}$$

where:

P_h = projected total population for stratum h on 23 December 2014

P_{14h} = population projection for stratum h on 1 July 2014

P_{15h} = population projection for stratum h on 1 July 2015

$t_{IOF} - t_{14}$ = number of days between 1 July 2014 and 23 December 2014 (that is, 175 days)

$t_{15} - t_{14}$ = number of days between 1 July 2014 and 1 July 2015 (that is, 365 days)

Table 3 presents the INE population projections by province, urban and rural stratum, for 1 July 2014 and 1 July 2015, and the corresponding interpolated population estimates for 23 December 2014.

Table 3. Mozambique Population Projections by Province, Urban and Rural Stratum for 2014 and 2015, and Interpolated Population for Mid-Point of IOF Data Collection Period for the Second Quarter

	2014	2015	IOF - Q2
Province and Stratum	1 July	1 July	23-12-14
Niassa Urban	372,176	388,202	379,775
Niassa Rural	1,221,307	1,268,704	1,243,806
Cabo Delgado Urban	444,864	463,038	453,487
Cabo Delgado Rural	1,417,221	1,430,118	1,423,390
Nampula Urban	1,549,414	1,615,298	1,580,660
Nampula Rural	3,338,425	3,393,495	3,364,716
Zambézia Urban	958,355	1,008,281	981,976
Zambézia Rural	3,724,080	3,794,084	3,757,481
Tete Urban	327,752	341,385	334,219
Tete Rural	2,090,829	2,176,059	2,131,268
Manica Urban	447,430	460,597	453,695
Manica Rural	1,418,871	1,472,925	1,444,535
Sofala Urban	725,458	737,503	731,208
Sofala Rural	1,273,851	1,311,173	1,291,611
Inhambane Urban	349,499	359,253	354,142
Inhambane Rural	1,125,819	1,140,226	1,132,704
Gaza Urban	358,546	365,350	361,792
Gaza Rural	1,033,526	1,051,460	1,042,086
Maputo Province Urban	1,145,642	1,200,866	1,171,795
Maputo Province Rural	492,989	508,192	500,221
Maputo City	1,225,868	1,241,702	1,233,434
Mozambique	25,041,922	25,727,911	25,368,001

Table 4 shows the population projections for the mid-point of the IOF data collection period for the second quarter, the IOF weighted estimates of total population by stratum based on the adjusted design weights, and the corresponding weight adjustment factor for the sample household weights in each stratum. It can be seen in Table 4 that the weight adjustment factors vary from 0.9852 for Cabo Delgado Rural to 1.5460 for Maputo Province Urban.

Table 4.   Mozambique Population Projections and IOF Weighted Estimates of Total Population for Second Quarter by Province, Urban and Rural Stratum, and Corresponding Weight Adjustment Factors

Province and Stratum	Projected Population 23-12-14	Weighted Population IOF, Second Quarter	Weight Adjustment Factor
Niassa Urban	379,775	255,172	1.4883
Niassa Rural	1,243,806	1,005,485	1.2370
Cabo Delgado Urban	453,487	353,239	1.2838
Cabo Delgado Rural	1,423,390	1,444,843	0.9852
Nampula Urban	1,580,660	1,118,842	1.4128
Nampula Rural	3,364,716	3,091,558	1.0884
Zambézia Urban	981,976	651,380	1.5075
Zambézia Rural	3,757,481	3,290,126	1.1420
Tete Urban	334,219	228,912	1.4600
Tete Rural	2,131,268	1,628,936	1.3084
Manica Urban	453,695	356,367	1.2731
Manica Rural	1,444,535	1,141,585	1.2654
Sofala Urban	731,208	704,927	1.0373
Sofala Rural	1,291,611	1,209,045	1.0683
Inhambane Urban	354,142	281,369	1.2586
Inhambane Rural	1,132,704	962,979	1.1762
Gaza Urban	361,792	281,592	1.2848
Gaza Rural	1,042,086	911,631	1.1431
Maputo Province Urban	1,171,795	757,940	1.5460
Maputo Province Rural	500,221	384,522	1.3009
Maputo City	1,233,434	1,055,780	1.1683

Megill worked closely with Basílio Cubula in adapting the IOF first quarter weighting spreadsheet for calculating the weights for the second quarter. They also worked together in obtaining the population projections. As soon as the household-level summary file from the final IOF data for the second quarter became available the week after Megill's mission, he used this information to complete the weighting spreadsheet for calculating the cross-sectional weights for the IOF second quarter. These weights were provided to INE and the Scanstat consultants in both SPSS and Excel formats. The Excel spreadsheets used for calculating the final weights and the population projections were sent to Basílio Cubula at INE.

2.2.2. Calculation of Cross-Sectional Weights for the First Six Months of IOF Data

The weights for the individual first and second quarters of IOF cross-sectional data were calculated independently based on the EAs and households enumerated each quarter, although the weighting procedures were consistent. Each survey was treated as a cross-sectional survey for the corresponding quarter, and the weights were adjusted at the

stratum level based on population projections for the mid-point of the corresponding quarter. INE would also like to tabulate indicators that represent the 6-month period corresponding to the first two quarters of IOF, and once the data for all quarters of IOF are available, estimates of annual indicators will be produced. Each of these data sets will require a different set of weights that is based on the sample included in the corresponding analysis.

When the IOF data for the first and second quarters are combined as a 6-month cross-sectional survey, it will simply be necessary to append the corresponding final data sets for the two quarters. However, first it will be necessary to include an additional variable in each data set, corresponding to the quarter (1 or 2). Therefore in the combined data set each household would be uniquely identified by the *Trimestre* (quarter 1 or 2), ID06 (Cluster) and ID07 (household number). The weights will be merged in the data file based on matching the *Trimestre* and ID06 variables.

The weighted cross-sectional data for each individual quarter represents all of Mozambique. Therefore when the data for two quarters are combined, it is necessary to divide the final quarterly weights by 2 so that the sum of the combined weights will be an estimate of the total number of households in Mozambique. Since the final cross-sectional weights for each quarter have already been adjusted based on the population projections by stratum for the mid-point of that quarter, the weighted estimate of total population using the quarterly weights (divided by 2) for the combined quarters will be equal to the average population projections for the mid-point of the 6 month data collection period. After calculating the final cross-sectional weights for the second quarter of IOF, Megill also calculated the cross-sectional weights for the combined IOF data from the first two quarters.

2.2.3. Weighting Procedures for the IOF Panel Data

The population represented by the panel corresponds to the households that are included in the first quarter and do not move during the period of data collection for the remaining quarters. In order to conduct a longitudinal analysis of the data for the first two quarters of IOF (and later for all quarters) using the data for a panel of households, it is necessary to link the data from the different quarters for each sample household. In this case it is necessary to identify which sample households will be used for the analysis. For example, if only the households with completed interviews in all quarters will be used for the panel analysis, then the weights would be calculated for this set of households. Since the panel was established in the first quarter of IOF, the population projections that would be used for adjusting the weights would be the same as those used for the IOF weights for the first quarter.

In the case of the panel analysis for the first and second quarters, panel weights were calculated for the households that had completed interviews for both quarters. Therefore it was first necessary to match the IOF household data files for the first and second quarters in order to determine which sample households had completed interviews for both quarters. The weights were then calculated for this subset of sample households. Given that only 1,175 sample EAs were enumerated in the second quarter, the panel data will also be limited to these sample EAs. Within these EAs a total of 10,186 households had completed interviews for both the first and second quarters.

The household-based population represented by the panel data corresponds to the frame for the first quarter. In order to ensure that the panel weights are consistent with

those of the cross-sectional weights, similar procedures were used for calculating the weights, including the weight adjustment based on population projections. The same formula used for calculating the second quarter cross-sectional IOF weights was used for calculating the panel weights for the first two quarters. The number of households in each sample EA corresponds to the number of households with completed interviews for both quarters, which generally is the same as the number of completed households in the second quarter. However, in this case the weighted total population for each stratum used for calculating the weight adjustment factor would be based on the number of persons in each household from the first quarter. The population projection for each stratum is also the same as that used for the adjustment of the first quarter cross-sectional weights. The weights were calculated in this way for the panel data for the first two quarters. Table 5 shows the population projections for the mid-point of the first quarter data collection, the weighted total population from the database of sample households with completed interviews for both the first and second quarters, and the corresponding panel weight adjustment factor for each stratum.

Table 5.   Mozambique Population Projections and Weighted Estimates of Total Population from IOF Panel Data by Province, Urban and Rural Stratum, and Corresponding Panel Weight Adjustment Factors

Province and Stratum	Projected Population 23-09-14	Weighted Population IOF, Panel First Quarter	Panel Weight Adjustment Factor
Niassa Urban	375,805	275,573	1.3637
Niassa Rural	1,232,055	1,043,956	1.1802
Cabo Delgado Urban	448,982	388,530	1.1556
Cabo Delgado Rural	1,420,179	1,481,545	0.9586
Nampula Urban	1,564,334	1,190,961	1.3135
Nampula Rural	3,351,019	3,273,440	1.0237
Zambézia Urban	969,621	683,582	1.4184
Zambézia Rural	3,740,075	3,418,497	1.0941
Tete Urban	330,840	224,829	1.4715
Tete Rural	2,110,143	1,660,716	1.2706
Manica Urban	450,426	373,832	1.2049
Manica Rural	1,431,132	1,197,610	1.1950
Sofala Urban	728,212	739,076	0.9853
Sofala Rural	1,282,345	1,241,175	1.0332
Inhambane Urban	351,720	285,407	1.2323
Inhambane Rural	1,129,118	1,013,606	1.1140
Gaza Urban	360,101	298,819	1.2051
Gaza Rural	1,037,626	952,617	1.0892
Maputo Province Urban	1,158,122	787,058	1.4715
Maputo Province Rural	496,447	404,809	1.2264
Maputo City	1,229,494	1,097,857	1.1199

A similar procedure will be used for calculating the weights for the final panel data based on all quarters of IOF.

2.3. Calculation of Sampling Errors

The methodology for calculating sampling errors for estimates of key survey indicators from the IOF data was described in Megill's December 2014 Mission Report, which can be used as a reference. Since the final IOF database for each quarter with corresponding final weights is needed for tabulating the sampling errors, it was not possible to tabulate the sampling errors during this mission. During his third mission Megill will assist the INE staff in using the Complex Samples module of SPSS for tabulating sampling errors for selected indicators by quarter, as well for the combined data from all quarters. He will also provide training in the use of this software.

2.4. Capacity Building

Since Megill had to adjust the scope of work for this mission to resolve the IOF data problems for the second quarter, he did not have time to provide more formal training in sampling as described in the terms of reference. However, he provided a considerable amount of on-the-job training to the INE staff throughout this visit. Although he produced the final weights remotely, he had spent a considerable amount of time during the mission working closely with Basílio Cubula, the main INE Sampling Statistician, on constructing the template for the calculation of the weights, and obtaining the population projections for the second quarter by province, urban and rural strata, that were needed for adjusting the final weights. The IOF weighting procedures for the second quarter are similar to those for the first quarter, and Megill had worked closely with Cubula on that weighting application during his previous mission.

3. FINDINGS AND RECOMMENDATIONS

The main findings during this mission are discussed in the previous section, and the highlights are summarized here. The main problem that needs to be resolved is that the IOF data collection has currently been stopped because the budget has not been released due to political problems. This introduces a seasonal gap in the IOF data for 12 months, corresponding to a period of relatively high agricultural production. This issue has to be discussed further with the analysts who will be working on the poverty study and other types of analysis, to see if there are some modelling techniques to adjust for the missing data, perhaps using quarterly trends from the 2008 IOF data.

Another issue that needs to be addressed is related to the EAs that could not be enumerated in the second quarter, especially for Zambézia, where 50 sample EAs were not covered due to flooding. The distribution of the 50 missing EAs in Zambézia by district was examined. All of the 10 sample EAs in the district of Ile were missing, as well as all of the 4 EAs in Namarroi; these EAs were all rural. In Alto Molocue all the 8 rural sample EAs were missing, and in Chinde all the 4 rural sample EAs are missing. In addition to these districts, half or more of the rural EAs are missing in Lugela, Maganja da Costa and Mocuba. This missing geographic coverage should be noted in the analysis of the IOF data for the second quarter. The weights for the EAs enumerated in Zambézia in the second quarter were adjusted to take into account the missing sample EAs in each stratum, but the results would still be affected by a corresponding bias. One way to study the potential bias would be to use the IOF data for the first quarter, and remove the data for the same 50 sample EAs missing in the second quarter. Some key indicators can be tabulated from the first quarter IOF data for Zambézia with and without these 50 EAs, and the results can be compared to determine the potential level of the bias from the geographic gap in the data for the second quarter. This bias will also affect the panel analysis, which is based on the sample households with completed interviews for both quarters.

The issue of coding errors in the second quarter data described previously were mostly related to problems with the CAPI data entry application or the use of an inconsistent data dictionary. Anne Abelseth quickly identified the source of these problems, and she is correcting the CAPI application that will be used for the next quarter of IOF.

It should also be pointed out that the problems with the missing sampling information related to combining small sample EAs and sub-dividing large sample EAs prior to the listing operation that affected the weighting procedures for the first quarter also affect the weights for the other quarters. These problems and the resulting need to calculate approximate weights for IOF were discussed in Megill's December 2014 Mission Report.

APPENDIX 1. Persons Contacted

Instituto Nacional de Estatística (INE)

Dr. João Loureiro, INE President
Manuel Gaspar, INE Vice-President
Arão Balate, Director, *Direcção de Censos e Inquéritos*
Antônio Adriano, Deputy Director, *Direcção de Censos e Inquéritos*
Cristóvão Muahio, Chief, *Departamento de Metodologia e Amostragem*
Basílio Cubula, Sampling Statistician
Antônio, Programmer
Angelo, Programmer
Ramiro Mouzinho
Tomás Bernardo
Carlos Creva, former INE Sampling Statistician

Scanstat
Anne Abelseth, IT Consultant
Lars Carlsson, Resident Advisor
Lars Lundgren, Household Surveys Consultant