STATISTICS DENMARK

Statistisk sentralbyrå
Statistics Norway

Statistiska centralbyrån
Statistics Sweden

MZ:2015:08

Mission Report

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

From 10 to 31 October 2015

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

TABLES

1. INTRODUCTION AND TERMS OF REFERENCE

The *Instituto Nacional de Estatística* (INE) has completed the data collection for the *Inquérito sobre o Orçamento Familiar* (IOF) 2014/15, or Household Budget Survey (HBS), in a nationally-representative sample over the 12-month period from August 2014 to August 2015. However, the data collection did not take place during the third quarter because of political and budgetary issues. This resulted in a 3-month gap in the annual data. The original sample consisted of 11,592 households in 1,236 sample census enumeration areas (EAs). 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 IOF, designed to obtain income and expenditure data for all four quarters to represent seasonality. One of the objectives of the IOF is to obtain measures of poverty and other socioeconomic indicators, and to provide information on total 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 a period of one year.

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. Because of administrative and budgetary problems the IOF data collection stopped for the third quarter, and then resumed from 19 May to 14 August for the fourth quarter. Following the data collection and processing for the first and second quarters of IOF, David Megill, the Scanstat Sampling Consultant, assisted INE with calculating the weights for the IOF data each quarter. He worked closely with his counterpart, Basílio Cubula, the INE Statistician responsible for sampling. Megill also worked closely with the other Scanstat short-term consultants in reviewing the data quality and producing preliminary results. The main purpose of Megill's third mission in October 2015 was to finalize the weights for the fourth quarter, for the annual cross-sectional IOF data, and for the panel of sample households that were interviewed in all three quarters. During this mission Megill overlapped with the other Scanstat consultants, Lars Lundgren and Anne Abelseth.

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

Third Mission starting on 12 of October 2015

- **Objective:** During the third mission the Sampling Consultant will follow up the findings and recommendations from the previous visits, and review the panel data from all four quarters of INCAF/IOF.
- **Activities:** The weighting procedures for the INCAF/IOF results for the full year will be finalized as well as the weights for the data from each individual quarter. Sampling errors and design effects will be tabulated for key survey estimates from the INCAF/IOF data for the full year as well as for the individual four quarters. Based on these results, the Sampling Consultant will assist INE with a final evaluation of the panel survey methodology, sample rotation scheme and continuous survey methodology. The estimation of quarterly trends in the unemployment and labor force characteristics based on the panel methodology will be reviewed, as well as the estimation of the components of household income and expenditure. A second line of activity is to review and propose a solution of the

sampling aspects of a possible integration between INCAF and the annual agricultural survey IAI.

–
– **Expected outputs:** The findings and recommendations will be presented in a final seminar on the INCAF/IOF methodology.
– **Reporting:** The last deliverable will be a final report on the evaluation of the INCAF/IOF sampling methodology and recommendations for future improvements.

Given that some aspects of IOF data editing had not been completed by the end of this mission, the calculation of the final weights required more time than expected. The calculation of the IOF weights for the consumption modules will be completed remotely following this mission once the edited IOF data files for all three quarters become available.

This report includes detailed documentation of the weighting procedures for the cross-sectional IOF data from the fourth quarter and for the annual data from all three quarters, as well as the weighting procedures for the panel data from all 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 Arão Balate, Basílio Cubula, 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, and the agenda for this visit. Of the original 1,236 sample EAs selected for IOF, 1,233 were covered during the first quarter, only 1,175 EAs were enumerated during the second quarter (because of major flooding in Zambézia), and 1,225 EAs were covered during the fourth quarter. Anne Abelseth began working with INE two weeks prior to Megill's arrival, so she was able to send him summary files from the IOF data for the fourth quarter before this mission. Megill identified 8 sample clusters that were missing in addition to the 3 clusters that were not included in the panel from the first quarter. Then INE verified that these 8 sample clusters could not be enumerated in the fourth quarter, so the data file was considered to be complete. During this mission INE was still conducting some edits for the IOF data from the fourth quarter, but by the second week the list of households with completed interviews had been finalized for all quarters. The basic weights were based on the distribution of the households with completed interviews by cluster. Since some of the households were missing expenditure and auto-consumption data, a separate set of weights will be calculated for the quarterly and panel consumption data. The calculation of the different sets of weights is described in the next sections.

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. The sample households selected in the first quarter were followed as a panel for the second and fourth quarters.

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 types: 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. The IOF cross-sectional data for all three quarters are combined vertically into one annual IOF cross-sectional data file, and a different set of weights is calculated for this file. Basically, the annual estimates of each indicator would be equal to the average of the indicators for all three quarters.

In the case of the panel survey data, we only include the data for households that have completed interviews in all quarters. Therefore it was necessary to match the identification codes of the households from all three quarters to identify the households that are included in the panel for the tabulations and analysis. Then a separate set of weights was calculated for the panel households. The methodology for calculating each different set of IOF weights is discussed below.

2.1. Calculation of IOF Cross-Sectional Weights for the Fourth 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. Later it was found that the expenditure data entered in the field was sometimes missing or the quality of the data

was poor, so INE decided that only the expenditure data entered in the central office would be used for the final edits and analysis.

Initially Anne Abelseth provided Megill with a summary file with the number of households with completed IOF questionnaires by cluster for the fourth quarter. He used this information to calculate the basic household sampling probabilities and weights. However, later it was found that this summary information was not consistent with the final data set for the fourth quarter, so he generated a new summary file using the SPSS software. This information was copied into a spreadsheet with the formulas for calculating the household sampling probabilities and weights for the fourth quarter of IOF.

Table 1 shows the distribution of the enumerated sample EAs and households for the fourth quarter of IOF by province, urban and rural stratum.

Table 1. Distribution of Enumerated Sample EAs and Households with Completed Interviews for the Fourth 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	316	64	501	96	817
Cabo Delgado	44	412	59	428	103	840
Nampula	60	592	104	761	164	1,353
Zambézia	52	478	117	811	169	1,289
Tete	40	422	68	508	108	930
Manica	40	414	56	421	96	835
Sofala	60	657	41	327	101	984
Inhambane	40	410	52	387	92	797
Gaza	40	425	48	363	88	788
Maputo Província	60	624	48	361	108	985
Maputo Cidade	100	967	0	0	100	967
Total	568	5,717	657	4,868	1,225	10,585

The weighting procedures for the fourth quarter of IOF data are similar to those used for the first and second quarters. 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 2014 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 basic weight for the cross-sectional data for the fourth quarter of IOF 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 fourth quarter of IOF

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

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

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 fourth 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 and fourth quarters of IOF, only the households with completed interviews in the first quarter were interviewed; since the non-interview households were not replaced for these quarters, the weights within each 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 fourth quarter by EA, in order to count 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 fourth quarter included a total of 1,225 EAs and 10,585 households with completed interviews, as shown in Table 1. A copy of the spreadsheet used for calculating the weights from the first quarter was adapted for the fourth quarter cross-sectional weights, since the information from the frame does not change. However, first it was necessary to identify and separate the 1,225 sample EAs that were enumerated in the fourth 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 using the population projections produced by INE, similar to the procedure that was used for the IOF weights for the first and second quarters. 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 mid-point of the IOF 2014/15 fourth quarter data collection, the preliminary weights were adjusted based on population projections.

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

$$A_{4h} = \frac{P_{4h}}{\sum_{i \in h} \sum_{j} \sum_{k} W''_{hij} \times p_{hijk}},$$

where:

A_{4h} = adjustment factor for the basic weights of the IOF sample households in stratum (province, urban/rural) h for the fourth quarter

P_{4h} = projected total population for stratum h for the mid-point of the data collection period for the fourth quarter of IOF, based on demographic analysis

W''_{hij} = adjusted fourth 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 fourth quarter

The denominator of the adjustment factor A_h is the estimated weighted total population in stratum h from the IOF data for the fourth 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:

$$W_{A4hij} = W''_{hij} \times A_{4h},$$

where:

W_{A4hij} = final adjusted weight for the cross-sectional sample households in the j-th sample EA of the i-th sample PSU in stratum h for the fourth quarter of IOF

After the adjustment factors were applied to the weights within each stratum, the final weighted survey estimates of total population by stratum were consistent with the corresponding population projections for the fourth 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 fourth quarter of IOF, the data collection was conducted between 19 May 2015 and 15 August 2015, so the mid-point was estimated as 2 July 2015. Since this is very close to the reference day for the 2015 population projections (1 July), we directly used the population projections for 2015 by stratum for adjusting the IOF weights for the fourth quarter. Table 2 shows the population projections for 1 July 2015, the IOF weighted estimates of total population by stratum based on the adjusted basic weights, and the corresponding weight adjustment factor for the sample household weights in each stratum for the fourth quarter of IOF. It can be seen in Table 2 that the weight adjustment factors vary from 0.9483 for Cabo Delgado Rural to 1.5278 for Tete Urban.

Table 2. Mozambique Population Projections and IOF Weighted Estimates of Total Population for Fourth Quarter by Province, Urban and Rural Stratum, and Corresponding Weight Adjustment Factors

Province and Stratum	Projected Population 1-7-15	Weighted Population IOF, Fourth Quarter	Weight Adjustment Factor
Niassa Urban	388,202	265,416	1.4626
Niassa Rural	1,268,704	1,049,673	1.2087
Cabo Delgado Urban	463,038	379,342	1.2206
Cabo Delgado Rural	1,430,118	1,508,036	0.9483
Nampula Urban	1,615,298	1,168,718	1.3821
Nampula Rural	3,393,495	3,242,281	1.0466
Zambézia Urban	1,008,281	709,313	1.4215
Zambézia Rural	3,794,084	3,437,419	1.1038
Tete Urban	341,385	223,448	1.5278
Tete Rural	2,176,059	1,667,836	1.3047
Manica Urban	460,597	372,051	1.2380
Manica Rural	1,472,925	1,196,026	1.2315
Sofala Urban	737,503	735,220	1.0031
Sofala Rural	1,311,173	1,236,481	1.0604
Inhambane Urban	359,253	285,167	1.2598
Inhambane Rural	1,140,226	1,035,397	1.1012
Gaza Urban	365,350	298,353	1.2246
Gaza Rural	1,051,460	954,369	1.1017
Maputo Province Urban	1,200,866	786,440	1.5270
Maputo Province Rural	508,192	402,828	1.2616
Maputo City	1,241,702	1,100,221	1.1286

Megill worked closely with Basílio Cubula in preparing the IOF weighting spreadsheet for calculating the weights for the fourth quarter. They also worked together in obtaining the population projections. These weights were provided to INE and the Scanstat consultants. The Excel spreadsheets used for calculating the final weights and the population projections were shared with Basílio Cubula at INE.

2.2. Calculation of Weights for IOF Annual Cross-Sectional Data

Once the cross-sectional weights were calculated for the fourth quarter of IOF, we could use the corresponding weights for all three quarters in order to calculate the cross-sectional annual weights. The weights for each quarter are used with the corresponding IOF quarterly data to represent Mozambique at the national and provincial levels for that reference period. Compiling the annual cross-sectional data involves combining (concatenating) the individual data files for all three quarters (with identical formats) into a single data file in a vertical manner. In this case the total

number of records in the cross-sectional annual IOF data file would be the sum of the number of records in the files for all three quarters. Since the IOF data from each quarter represents one third of the annual estimate for each indicator (such as the unemployment rate), the annual estimate would be the equivalent of the average of the estimates for the three quarters. For this reason the annual weight for each sample household in the combined cross-sectional data file would simply be equal to the final quarterly weight divided by 3. Since the cross-sectional weights for each quarter were adjusted based on the population projections for the mid-point of the corresponding quarter, the weighted estimate of the total population based on the annual data would be consistent with the population projections close to the mid-point of the IOF data collection for the 12-month period. For this reason it was not necessary to have a separate adjustment of the IOF cross-sectional annual weights based on the population projections. The calculation of the IOF annual weights involved returning to the weighting spreadsheet for each quarter, dividing the final quarterly weight by 3, and then compiling an SPSS database with the identification of all the IOF sample clusters by quarter, and the corresponding final quarterly and annual weights. It should be noted that the weight for the sample households in each cluster will vary by quarter. Therefore it is necessary to use both the quarter code (trimestre) and IOF cluster code (ID06) as keys to merge the cross-sectional annual weights in the IOF annual data file. An SPSS database and Excel spreadsheet with the final IOF cross-sectional annual weights by quarter and cluster were shared with the IOF analysts and attached to the combined IOF database with the employment data for all three quarters. Both the quarterly and annual cross-sectional weights were attached to the combined annual employment data file, so that the same database can be used for tabulating the quarterly and annual employment tables. A different set of IOF weights was calculated for the annual panel data, as described in the next section.

2.3. Calculation of Weights for IOF Annual Panel Data

For a panel survey, the sample households in the first quarter are enumerated each following quarter so that the household 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, and the panel weights will be different from the cross-sectional weights for each quarter.

The INE analysts and some other data users will be using the IOF panel data for all three quarters for some types of analysis and tabulations. In this case the analysis is limited to sample households that have completed IOF questionnaires for all three quarters. Ultimately it is possible to link the data from all three quarters for each individual household and household member to conduct a micro-level longitudinal analysis to follow the employment trends for individuals, for example. However, it is also possible to tabulate the panel data using a vertically concatenated database of the employment data from each quarter exclusively for the panel households. In this case the panel is only identified at the household level, so the panel consists of all the persons in each panel household that are included in the database for each quarter. That is, the person records in each sample panel household for all three quarters are included in the annual panel data file.

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. These households are identified in the IOF data file as a ”new” household. Even if the corresponding household identification appears in all three quarters, it was decided to exclude the ”new” sample households in the second and fourth quarters from the panel since they have different persons from those interviewed in the original sample household from the first quarter.

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 corresponding 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.

The first step in developing the weighting application for the IOF annual panel weights involved obtaining the database of households with completed interviews from each of the three quarters. After excluding the ”new” households from the second and fourth quarters, the unique household identification numbers (ID06 and ID07) was matched for all three quarters to identify the households with completed interviews in all quarters. These households were identified as the final annual panel for the longitudinal analysis, and they were assigned a code of 1 for a new panel variable. A separate file was generated with these panel households, and the data were aggregated to determine the total number of panel households in each sample EA. The final distribution of the EAs with panel households by province, urban and rural stratum was also tabulated. The final set of panel households are found in 1,168 sample EAs. A total of 1,175 EAs were enumerated in the second quarter of IOF, so apparently 7 of these EAs did not have any sample households with completed interviews in the fourth quarter. Table 3 shows the final distribution of the sample EAs and panel households by province, urban and rural stratum.

Table 3. Distribution of Enumerated Sample EAs and Panel Households with Completed Interviews for All Quarters (1, 2 and 4) of IOF 2014/15, Used for the Panel Survey Analysis, 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	277	63	418	95	695
Cabo Delgado	44	354	56	377	100	731
Nampula	60	525	104	715	164	1,240
Zambézia	43	357	77	508	120	865
Tete	39	366	65	459	104	825
Manica	40	382	56	408	96	790
Sofala	60	576	41	316	101	892
Inhambane	40	379	52	371	92	750
Gaza	40	407	48	352	88	759
Maputo Província	60	601	48	340	108	941
Maputo Cidade	100	854	0	0	100	854

Total	558	5,078	610	4,264	1,168	9,342

The steps involved in calculating the weights for the final set of panel households are similar to those described previously for the fourth quarter cross-sectional weights. A similar weighting spreadsheet was developed for calculating the panel weights, limited to the 1,168 sample EAs that are included in the panel IOF data. In this case it was necessary to update the column for the number of sample EAs in each stratum to reflect the distribution of the sample EAs in Table 3. The column for the number of sample households in each EA was also changed to include only the panel households. The same formulas were used for calculating the basic panel weights. A similar weight adjustment based on the INE population projections by stratum was also used for the panel weights. However, in this case the reference date for the population projections was based on the mid-point of the IOF data collection period, which was estimated to be 10 February 2015. Since the INE population projection tables are only available for 1 July 2014 and 1 July 2015, it was necessary to make an interpolation based on an exponential population growth rate to estimate the projected total population by province, urban and rural stratum for 10 February 2015. The following formula was used:

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

where:

P_h = projected total population for stratum h on 10 February 2015 (mid-point of IOF data collection

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 10 February 2015 (that is, 224 days)

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

After we tabulated the weighted total population by stratum using the basic panel weights and calculated the projected total population on 10 February 2015 by stratum, we used the same weight adjustment procedures described for the fourth quarter cross-sectional weights. The weight adjustment factor for each stratum is simply the ratio of the projected total population for the stratum divided by the corresponding preliminary weighted total population from the IOF data. Table 4 presents the INE population projections by province, urban and rural stratum, for 1 July 2014 and 1 July 2015, the corresponding interpolated population estimates for 10 February 2015, the preliminary weighted total population from the IOF data, and the weight adjustment factor for each stratum.

Table 4. Mozambique Population Projections by Province, Urban and Rural Stratum for 2014 and 2015, Interpolated Population for Mid-Point of IOF Data Collection Period for the Year, Preliminary Weighted Total Population from IOF Panel Data, and Corresponding Panel Weight Adjustment Factors

Province and Stratum	2014 01-07-14	2015 01-07-15	IOF - Annual 10-02-15	Weighted Population IOF, Annual Panel	Panel Weight Adjustment Factor
Niassa Urban	372,176	388,202	381,931	265,526	1.4384
Niassa Rural	1,221,307	1,268,704	1,250,180	1,034,006	1.2091
Cabo Delgado Urban	444,864	463,038	455,931	384,895	1.1846
Cabo Delgado Rural	1,417,221	1,430,118	1,425,122	1,461,072	0.9754
Nampula Urban	1,549,414	1,615,298	1,589,521	1,167,928	1.3610
Nampula Rural	3,338,425	3,393,495	3,372,115	3,216,046	1.0485
Zambézia Urban	958,355	1,008,281	988,693	695,696	1.4212
Zambézia Rural	3,724,080	3,794,084	3,766,887	3,445,041	1.0934
Tete Urban	327,752	341,385	336,053	229,644	1.4634
Tete Rural	2,090,829	2,176,059	2,142,730	1,654,232	1.2953
Manica Urban	447,430	460,597	455,465	375,524	1.2129
Manica Rural	1,418,871	1,472,925	1,451,804	1,206,197	1.2036
Sofala Urban	725,458	737,503	732,826	737,673	0.9934
Sofala Rural	1,273,851	1,311,173	1,296,628	1,238,349	1.0471
Inhambane Urban	349,499	359,253	355,453	287,441	1.2366
Inhambane Rural	1,125,819	1,140,226	1,134,639	1,018,469	1.1141
Gaza Urban	358,546	365,350	362,706	296,880	1.2217
Gaza Rural	1,033,526	1,051,460	1,044,495	949,060	1.1006
Maputo Province Urban	1,145,642	1,200,866	1,179,224	786,004	1.5003
Maputo Province Rural	492,989	508,192	502,264	408,410	1.2298
Maputo City	1,225,868	1,241,702	1,235,561	1,108,372	1.1148
Mozambique	25,041,922	25,727,911	25,460,230	21,966,465	

2.4. Weighting Procedures for the IOF Consumption Data

It would be possible to use the IOF cross-sectional and panel weights specified in the previous sections for all the IOF data, including the expenditure and consumption data. However, one problem is that not all of the sample households have complete data for the daily and monthly expenditures and auto-consumption. In order to estimate total food and non-food expenditures, and other consumption aggregates needed to determine the poverty indicators, some of the households will have missing data and will therefore need to be dropped from the poverty analysis. Conceptually the weights should be calculated for the specific set of sample households that will be included in the data analysis. Therefore if a considerable number of sample households will not have sufficient consumption data and will therefore be dropped from the poverty analysis, the use of the regular IOF cross-sectional and panel weights based on all households will result in biased estimates. For this reason it will be necessary to calculate a different set of cross-sectional and panel weights for the IOF consumption

data, once the final set of sample households that will be included in the consumption and poverty analysis has been determined.

At the end of this mission the INE staff were still working on the editing of the IOF expenditure and auto-consumption data, and the clean IOF data set may not be available until the end of November. For this reason Megill has agreed to assist remotely with the calculation of the weights for the final cross-sectional and panel consumption data later once the edited IOF data files are finalized.

The calculation of the cross-sectional and panel weights for the consumption data will be similar to the corresponding weights used for all the households with completed data, including the adjustment of the weights using population projections. The only difference is that the number of households with consumption data by EA may be less than the corresponding number of households used for calculating the original household weights. It will be necessary to calculate separate cross-sectional weights for the consumption data from each quarter, and then calculate the annual cross-sectional weights in the same way specified previously for the original households weights. The adjustment of the weights based on population projections will also be done in the same way for each quarter, using the same reference dates for the population projections. The annual cross-sectional and panel weights will also be calculated in the same way, so reference can be made to that methodology described previously.

2.5. Calculation of Sampling Errors

For the analysis of the IOF annual cross-sectional and panel estimates of labor force and unemployment indicators, the SPSS files with the employment data for all three quarters were vertically combined (concatenated). First it was necessary to ensure that the format and variable names for the three quarterly data files were consistent. Once the household cross-sectional and panel weights for the IOF annual data were calculated, these weights were merged in the combined annual employment data file, as well as an indicator variable that identifies the panel households. Therefore this file can be used to tabulate both cross-sectional and panel estimates, once the appropriate weights are specified.

Megill worked with the INE staff in using the SPSS Complex Samples software with the IOF annual cross-sectional and panel employment data file to tabulate estimates, sampling errors and design effects for the unemployment rate (for both the ILO and national definitions), and the labor force participation rate. This SPSS program uses a linearized Taylor series variance estimator for calculating the standard error for each indicator, which is the same variance estimator used by the Stata software.

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. The first step in using the Complex Samples module is to create a sample design specifications file (csplan), where we specify the stratum, cluster and weight variables. Megill worked with the INE staff in developing the sampling error application using the SPSS Complex Samples module, and provided them with brief training.

The results of the Complex Samples tabulation of sampling errors and design effects for the unemployment rate (both ILO and national definitions) and the labor force participation rate by domain are shown in the tables of Annex I. For each indicator and

category of a classification variable, the Complex Samples output tables include the value of the estimate, the standard error, coefficient of variation, the 95% confidence interval, the design effect and the number of observations. Each indicator was tabulated at the national level, by quarter, gender, urban and rural domains and province. Separate tables were produced for the IOF annual cross-sectional and panel survey data. It can be seen in these tables that the estimates from the annual cross-sectional and panel survey data are fairly close, given that most of the sample households in the cross-sectional data are also included in the panel data.

The tables in Annex A indicate that most of the the IOF estimates have a good level of precision even at the provincial level, given the relatively large sample size. In the case of the unemployment rate based on the national definition, the estimate for the second quarter (15.8% based on the cross-sectional data) is significantly lower than the corresponding estimate from the first quarter (25.2%). The design effects for the annual estimates are generally considerably higher than the corresponding quarterly estimates, given that the annual estimates have a much higher clustering effect from the three interviews in the same households, whereas within each quarter the households are only interviewed once.

The INE staff can use this SPSS Complex Samples application as a model for tabulating the sampling errors and confidence intervals for other key indicators from the IOF data, including the consumption aggregates once the corresponding final data files become available.

2.6. Capacity Building

Since Megill had to adjust the scope of work for this mission due to the delayed editing of the IOF data, 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. He spent a considerable amount of time during the mission working closely with Basílio Cubula, the main INE Sampling Statistician, on adapting the spreadsheet for the calculation of the weights for the cross-sectional and panel data, and obtaining the population projections for the fourth quarter by province, urban and rural strata, that were needed for adjusting the final weights. The IOF weighting procedures for the fourth quarter are similar to those for the previous quarters, and Megill had also worked closely with Cubula on those weighting applications during his previous missions.

During this mission Megill also provided some training to the INE staff in the use of the Complex Samples module of the SPSS software for tabulating standard errors, design effects and other measures of precision for estimates of key IOF employment indicators, as described in the previous section.

Another important source of capacity building is comprehensive documentation of all the IOF sampling and estimation methodology. This type of documentation has been provided with each mission report. These documents can be used for future reference to summarize the methodology for IOF reports, and to plan for future surveys.

2.7. Considerations for Combining the IAI with INCAF

The Integrated Agricultural Survey (*Inquérito Agrícola Integrado*, IAI) is conducted each year by the Ministry of Agriculture to produce estimates of total crop and livestock production, as well as socioeconomic characteristics of the farm households. This survey is based on the integration of two previous agricultural surveys that were conducted independently. The *Aviso Prévio* (Crop Forecasting Survey) was designed to produce an early forecast of the level of crop production. The *Trabalho do Inquérito Agrícola* (TIA) was designed to provide more accurate post-harvest estimates of crop production as well as livestock data, and more detailed characteristics of the farm households including consumption.

The *Inquérito Contínuo de Agregados Familiares* (INCAF) was designed as a continuous household survey for providing quarterly employment statistics, and can include modules on different topics each quarter. Between August 2014 and August 2015 the IOF was combined with the INCAF to provide 12 months of income and expenditure data for the analysis of poverty, and to provide information for national accounts, in addition to collecting the labor force and employment data covered by INCAF.

Both the IAI and INCAF are designed to be conducted at the national level each year, with a representative sample of households at the provincial level. Therefore the Scanstat project would like to examine the feasibility of combining the data collection for these two surveys. It is necessary to consider both the logistical and sampling issues that would be involved in this type of integration of two different household surveys.

First, it is important to compare the sampling frames and current samples for the IAI and INCAF. The sampling frame for the IAI is currently based on the sample enumeration areas (EAs) selected for the Census of Agriculture and Livestock (*Censo Agro-Pecuário*, CAP II), which is used as a master sample for the agricultural surveys. The CAP sampling frame excludes EAs with less than 15 agricultural households, so most of the urban EAs are excluded from the IAI sampling frame. On the other hand, the INCAF is based on INE's master sample for national household surveys, which covers the households in all of the urban and rural EAs. Therefore the IAI sampling frame can be considered to be a subset of the INCAF sampling frame. This provides more of a challenge for combining the two surveys, although it does not limit the potential for integrating the common part of the sampling frames for the two surveys. The stratification of the sampling frame of EAs for the two surveys is reasonably compatible, since both frames include implicit stratification of the rural EAs by agro-climatic zones. However, the IAI sampling methodology involves an additional second stage stratification of the households listed in each sample EA by farm size, so that the medium and large farms identified in the listing (based on farm size and the number of animals) can be included in the sample with certainty at the second sampling stage. Therefore one of the more challenging aspects of combining the INCAF and IAI samples would be the final stage of selecting the households in the sample EAs. In the case of INCAF the households in each EA are selected with equal probability to improve the efficiency of the sample design. The IAI has a separate frame of large farms which may not be covered separately for the INCAF, and the sampling frame for the 2015 IAI has a special stratification of EAs with a high concentration of cattle, in order to improve the level of precision for the estimates of total cattle production.

Another aspect of the IAI methodology that will make it more difficult to coordinate the data collection for the two surveys is that the IAI includes a data collection phase for crop forecasting that is scheduled based on the agricultural calendar. This involves crop cutting prior to the harvest, so the schedule has to be carefully planned and followed.

The post-harvest component of the data collection is also very sensitive to the agricultural calendar. On the other hand, the data collection for INCAF involves returning to a nationally representative panel of sample households each quarter. Therefore the schedule of the data collection would have to be carefully coordinated in a combined survey.

Given that the IAI and INCAF are carried out by two different government agencies, the integration of these surveys would need to be supported politically at the highest level of each organization, and coordinated between the two institutions. Given that INE is responsible for the National Statistical System, it would need to take the lead in any effort to integrate the INCAF and IAI. The human and financial resources available for each survey would need to be combined and centrally managed in order to integrate the data collection for the two surveys. If there is a political will for combining the data collection for the IAI and INCAF, then it should be possible to overcome the individual challenges described here.

Both the technical aspects and logistics of the survey operations would have to be coordinated based on the timing requirements for both surveys. Although most of the rural sample for INCAF can overlap at least with the IAI sample of small farms, part of the sample may only be included in INCAF or IAI. For example, most of the urban sample for INCAF would be out of scope for IAI, and the frame of large farms selected with certainty for IAI may not be included in the INCAF sample.

Another important issue that would need to be addressed is the integration of the questionnaires for the INCAF and IAI. It will be necessary to determine the IAI questions that would need to be included in the INCAF questionnaire in the corresponding quarters. The crop forecasting questions would need to be included for the appropriate quarter based on the agricultural calendar, and the post-harvest crop production questions would need to be included in the quarter following the main crop harvest. As indicated previously, the timing of the crop forecasting and post-harvest questions is very critical to the objectives of the IAI.

In conclusion, although it would be technically possible to combine much of the data collection for the IAI and the INCAF over each period of 12 months, it would require a high level of coordination between INE and the Ministry of Agriculture, and a strong political will at the highest levels. Sufficient resources would need to be provided for conducting the integrated survey, which should be available according to a pre-determined timeline. In the past both INE and the Ministry of Agriculture have had periods of gaps in the release of funds when they are needed for a continuous survey program, so this issue needs to be resolved.

3. FINDINGS AND RECOMMENDATIONS FROM ALL CONSULTANT MISSIONS FOR IOF

The main findings during the previous visits were discussed in the corresponding mission reports. However, these issues are summarized here, together with the findings from this last mission, and corresponding recommendations.

Although the data collection for the three quarters of IOF 2014/15 was successful and the survey data appear to have reasonable quality, there were some important lessons learned that affected the survey in all quarters. Sampling information related to

combining small sample EAs and sub-dividing large sample EAs prior to the listing operation appears to have been lost. This information would be needed to calculate the exact probabilities and corresponding weights for the IOF sample households. Since this information was not available, it was necessary to calculate approximate weights which were then adjusted based on the population projections by province, urban and rural stratum. Since the IOF is based on a panel of households that are enumerated each quarter, it was necessary to use the same approximate weighting procedures for all quarters. For future household surveys it is recommended that the information from each sampling stage should be carefully recorded and maintained for the calculation of the probabilities of selection and corresponding design weights.

Conceptually, a complete listing of households in the sample EAs reflects the overall average growth in the number of households across all the sample EAs, so the weighted estimates of the total population would also show a corresponding increase. Therefore the design weights depend on the updating of the sampling frame based on the listing, and if the listing for some sample EAs is not complete, this will lead to a downward bias in the weighted population estimates from the survey data. It is important to note that it is ideal to rely on a high-quality updated listing of households in each sample EA and weights based on the sampling probabilities to reflect the differential population growth by province, urban and rural stratum. Although it was too late to correct this for IOF 2014/15, this is an important lesson learned for improving future surveys. The population projections are based on the growth rates between the last two censuses and general demographic assumptions, so they do not always accurately reflect the actual differential growth rates by urban and rural stratum within each province. For this reason it is not good to always rely on the population projections for adjusting the probability-based weights.

The main problem that affected the quality of the IOF data is that the data collection was stopped during the third quarter, mostly because of political issues that affected the release of funding to continue the IOF fieldwork. 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 also

affects the panel analysis, which is based on the sample households with completed interviews for all three quarters.

There were problems related to coding errors in the second quarter data, 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 also corrected the CAPI application that will was used for the final quarter of IOF. However, these problems resulted in a loss of staff time and corresponding delays.

The data capture for the IOF consumption data has to be improved for future surveys. Although the original survey methodology involved a CAPI application in the field for the consumption data, a paper questionnaire was introduced to ensure that there was a back-up so that the consumption data could be entered again in the office. Ideally a CAPI application should be designed to catch any errors in the field where they could be corrected, so that the consumption data would be fairly clean by the time they are sent to the office. However, there was a lack of quality and operational control for the data entry of the consumption data in the field, so there was missing and inconsistent information in the data that were received from the field. As a result, these data were not used, and instead it was necessary to rely on a single entry data keying operation in the office that required a considerable amount of editing. Delays in the data entry and editing of the IOF expenditure and auto-consumption data resulted in corresponding delays in the availbility of the consumption data for the analysis. Even with the problems found with the consumption data captured in the field, the editing of the consumption data from the first two quarters entered in the office could have been better coordinated so that all the editing rules could be finalized by the beginning of the fourth quarter.

Despite the various problems discussed in this section, a review of the IOF consumption and employment data for the three quarters by the consultants indicates that the corresponding quality appears to be fairly good. Also, the issues related to the seasonal and geographic gaps in the IOF data described here can be studied further and perhaps corresponding adjustments can be made by the analysts for some aspects of the poverty analysis and other results. Overall there were important lessons learned for future reference.

Regarding the future methodology for INCAF, this will depend on INE's long-term plans for their natonal household survey program. If they would like to produce quarterly labor force and unemployment statistics each year, planning for a continuous national household survey such as INCAF would be an effective approach. Modules on different subjects such as health and education could be included for particular quarters, since this would be a multi-purpose national household survey. The sample size and allocation should be studied further to ensure not only that accurate survey results can be published on a timely basis, but that it will be possible to ensure the sustainability of INCAF over time.

INE is planning to conduct the next Mozambique Census of Population and Housing in August 2017. Therefore in 2016 much of their staff and resources will be concentrated on preparing for this Census, including a Pilot Census in August 2016, and a follow-on pilot of the Post-Enumeration Survey (*Inquérito de Cobertura*). Therefore it is not clear whether they will plan to conduct the contiuous INCAF during the period prior to the 2017 Census. Following the Census INE plans to develop a new master sampling frame (*Amostra Mãe*), and this will provide an ideal opportunity to begin the new INCAF

series based on the new sampling frame, and improve the sampling methdology, as well as the quality and operational control procedures. This will improve the representativeness of the sample based on this new frame, and avoid the types of sampling issues discussed in this and previous mission reports.

APPENDIX 1. Persons Contacted

Instituto Nacional de Estatística (INE)

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

APPENDIX 2. Tables of estimates, sampling errors, coefficients of variation (CVs), 95% confidence intervals, design effects and number of observations for key indicators from annual 2014/15 Mozambique IOF

Table A1. Cross-sectional estimates of the unemployment rate (ILO definition) by domain, 2014/15 IOF annual data

Domain	Estimate (%)	SE (%)	CV	95% confidence interval		DEFF	No. observations (denominator)
				Lower (%)	Upper (%)		
Mozambique	6.1	0.22	0.037	5.6	6.5	5.8	66,704
Quarter							
First	7.4	0.28	0.038	6.9	8.0	2.7	24,526
Second	5.2	0.26	0.049	4.7	5.7	3.1	21,299
Fourth	5.6	0.27	0.049	5.0	6.1	3.0	20,879
Gender							
Male	5.2	0.22	0.043	4.7	5.6	3.2	31,466
Female	6.8	0.27	0.040	6.3	7.4	4.1	35,238
Residence							
Urban	15.7	0.61	0.039	14.5	16.9	5.8	36,119
Rural	1.8	0.14	0.082	1.5	2.0	5.5	30,585
Province							
Niassa	5.9	0.76	0.130	4.4	7.4	4.0	4,464
Cabo Delgado	4.3	0.64	0.150	3.0	5.5	5.3	5,467
Nampula	7.9	0.74	0.094	6.4	9.3	9.7	8,057
Zambézia	2.5	0.43	0.172	1.7	3.4	9.5	7,262
Tete	2.5	0.32	0.124	1.9	3.2	2.6	5,392
Manica	3.2	0.33	0.103	2.5	3.8	1.8	6,153
Sofala	5.8	0.68	0.118	4.5	7.2	4.5	6,507
Inhambane	4.2	0.44	0.104	3.3	5.0	1.8	4,722
Gaza	4.8	0.54	0.112	3.8	5.9	2.3	5,135
Maputo Província	12.7	1.14	0.090	10.5	15.0	5.6	6,508
Maputo Cidade	19.7	1.05	0.053	17.7	21.8	2.5	7,037

Table A2. Cross-sectional estimates of the unemployment rate (national definition) by domain, 2014/15 IOF annual data

Domain	Estimate (%)	SE (%)	CV	95% confidence interval		DEFF	No. observations (denominator)
				Lower (%)	Upper (%)		
Mozambique	20.7	0.46	0.022	19.8	21.6	8.6	66,704
Quarter							
First	25.2	0.66	0.026	23.9	26.5	5.3	24,526
Second	15.8	0.53	0.034	14.8	16.9	4.8	21,299
Fourth	21.1	0.67	0.032	19.8	22.5	5.8	20,879
Gender							
Male	19.4	0.48	0.025	18.4	20.3	4.7	31,466
Female	21.9	0.55	0.025	20.8	23.0	6.3	35,238
Residence							
Urban	29.4	0.69	0.023	28.1	30.8	4.7	36,119
Rural	16.8	0.58	0.035	15.7	18.0	11.3	30,585
Province							
Niassa	25.7	1.40	0.055	22.9	28.4	3.9	4,464
Cabo Delgado	16.1	1.37	0.085	13.4	18.8	7.3	5,467
Nampula	19.6	1.13	0.057	17.4	21.9	10.4	8,057
Zambézia	13.4	0.94	0.070	11.6	15.3	9.4	7,262
Tete	20.2	2.21	0.109	15.9	24.6	19.5	5,392
Manica	23.3	1.47	0.063	20.5	26.2	6.2	6,153
Sofala	18.1	1.58	0.088	15.0	21.2	9.0	6,507
Inhambane	21.0	1.35	0.064	18.4	23.7	4.3	4,722
Gaza	30.0	2.50	0.083	25.1	34.9	11.0	5,135
Maputo Província	31.4	1.33	0.042	28.8	34.0	3.9	6,508
Maputo Cidade	28.1	1.12	0.040	25.9	30.3	2.2	7,037

Table A3. Cross-sectional estimates of the labor force participation rate by domain, 2014/15 IOF annual data

Domain	Estimate (%)	SE (%)	CV	95% confidence interval		DEFF	No. observations (denominator)
				Lower (%)	Upper (%)		
Mozambique	85.0	0.34	0.004	84.3	85.6	7.7	83,190
Quarter							
First	82.5	0.48	0.006	81.5	83.4	4.7	31,263
Second	86.5	0.40	0.005	85.7	87.3	3.9	26,307
Fourth	86.1	0.42	0.005	85.3	87.0	3.8	25,620
Gender							
Male	85.0	0.41	0.005	84.2	85.8	5.1	38,675
Female	84.9	0.38	0.004	84.2	85.7	5.0	44,515
Residence							
Urban	75.0	0.59	0.008	73.8	76.1	5.5	49,236
Rural	90.4	0.43	0.005	89.5	91.2	11.5	33,954
Province							
Niassa	84.1	1.46	0.017	81.2	86.9	7.7	5,654
Cabo Delgado	86.4	1.36	0.016	83.7	89.0	10.0	6,813
Nampula	87.2	1.05	0.012	85.1	89.2	15.4	9,488
Zambézia	86.4	0.90	0.010	84.6	88.1	10.6	8,781
Tete	90.4	0.95	0.011	88.5	92.2	7.9	6,686
Manica	87.9	0.90	0.010	86.2	89.7	4.7	7,293
Sofala	83.5	0.95	0.011	81.6	85.4	4.4	8,308
Inhambane	80.7	1.02	0.013	78.7	82.7	3.4	6,214
Gaza	85.1	1.06	0.012	83.0	87.2	4.1	6,283
Maputo Província	81.4	0.72	0.009	80.0	82.8	2.1	7,989
Maputo Cidade	72.3	1.13	0.016	70.1	74.6	3.4	9,681

Table A4. Panel estimates of the unemployment rate (ILO definition) by domain, 2014/15 IOF annual data

Domain	Estimate (%)	SE (%)	CV	95% confidence interval		DEFF	No. observations (denominator)
				Lower (%)	Upper (%)		
Mozambique	6.0	0.23	0.038	5.6	6.5	5.4	58,884
Quarter							
First	7.3	0.29	0.040	6.7	7.8	2.5	20,598
Second	5.2	0.25	0.049	4.7	5.7	2.6	19,582
Fourth	5.6	0.28	0.050	5.0	6.1	2.8	18,704
Gender							
Male	5.2	0.23	0.044	4.7	5.6	2.9	27,584
Female	6.8	0.28	0.042	6.2	7.3	4.0	31,300
Residence							
Urban	15.6	0.62	0.040	14.4	16.8	5.4	32,023
Rural	1.7	0.16	0.091	1.4	2.1	6.0	26,861
Province							
Niassa	6.1	0.83	0.134	4.5	7.8	4.0	3,820
Cabo Delgado	4.3	0.64	0.151	3.0	5.5	4.7	4,573
Nampula	7.8	0.72	0.092	6.4	9.3	8.2	7,241
Zambézia	2.3	0.50	0.219	1.3	3.3	12.3	4,956
Tete	2.5	0.33	0.129	1.9	3.2	2.5	4,830
Manica	3.2	0.34	0.107	2.5	3.8	1.7	5,814
Sofala	5.6	0.70	0.124	4.2	7.0	4.3	6,006
Inhambane	4.4	0.45	0.102	3.5	5.3	1.6	4,392
Gaza	4.8	0.55	0.113	3.7	5.9	2.1	4,934
Maputo Província	12.7	1.14	0.090	10.4	14.9	5.1	6,161
Maputo Cidade	20.1	1.11	0.055	17.9	22.2	2.4	6,157

Table A5. Panel estimates of the unemployment rate (national definition) by domain, 2014/15 IOF annual data

Domain	Estimate (%)	SE (%)	CV	95% confidence interval		DEFF	No. observations (denominator)
				Lower (%)	Upper (%)		
Mozambique	20.7	0.48	0.023	19.8	21.7	8.3	58,884
Quarter							
First	25.3	0.69	0.027	24.0	26.7	5.2	20,598
Second	15.7	0.53	0.034	14.7	16.7	4.2	19,582
Fourth	21.0	0.69	0.033	19.7	22.4	5.4	18,704
Gender							
Male	19.5	0.51	0.026	18.5	20.5	4.5	27,584
Female	21.8	0.57	0.026	20.7	22.9	6.1	31,300
Residence							
Urban	29.4	0.71	0.024	28.0	30.8	4.5	32,023
Rural	16.8	0.62	0.037	15.6	18.1	11.1	26,861
Province							
Niassa	26.2	1.44	0.055	23.4	29.0	3.6	3,820
Cabo Delgado	16.2	1.38	0.085	13.5	18.9	6.4	4,573
Nampula	19.7	1.13	0.057	17.5	21.9	9.2	7,241
Zambézia	13.1	1.13	0.086	10.9	15.3	12.2	4,956
Tete	20.1	2.29	0.114	15.6	24.6	18.7	4,830
Manica	23.5	1.47	0.063	20.6	26.4	5.5	5,814
Sofala	17.9	1.59	0.089	14.7	21.0	8.1	6,006
Inhambane	20.8	1.26	0.061	18.3	23.3	3.3	4,392
Gaza	30.0	2.60	0.087	24.9	35.1	10.6	4,934
Maputo Província	31.5	1.34	0.043	28.9	34.1	3.6	6,161
Maputo Cidade	28.6	1.18	0.041	26.3	30.9	2.1	6,157

Table A6. Panel estimates of the labor force participation rate by domain, 2014/15 IOF annual data

Domain	Estimate (%)	SE (%)	CV	95% confidence interval		DEFF	No. observations (denominator)
				Lower (%)	Upper (%)		
Mozambique	84.9	0.37	0.004	84.1	85.6	8.0	73,338
Quarter							
First	82.5	0.52	0.006	81.5	83.5	4.8	26,208
Second	86.3	0.42	0.005	85.4	87.1	3.5	24,159
Fourth	86.1	0.44	0.005	85.3	87.0	3.8	22,971
Gender							
Male	84.8	0.45	0.005	83.9	85.7	5.5	33,916
Female	84.9	0.41	0.005	84.1	85.7	5.1	39,422
Residence							
Urban	74.9	0.62	0.008	73.7	76.1	5.4	43,542
Rural	90.3	0.48	0.005	89.3	91.2	12.6	29,796
Province							
Niassa	84.6	1.22	0.014	82.2	87.0	4.8	4,850
Cabo Delgado	85.9	1.62	0.019	82.8	89.1	12.2	5,699
Nampula	87.2	1.08	0.012	85.1	89.3	14.3	8,463
Zambézia	85.9	1.14	0.013	83.7	88.2	14.6	6,025
Tete	90.3	0.97	0.011	88.4	92.2	7.2	6,008
Manica	87.7	0.92	0.011	85.9	89.5	4.3	6,885
Sofala	83.3	1.01	0.012	81.3	85.3	4.3	7,646
Inhambane	80.6	1.13	0.014	78.4	82.8	3.6	5,761
Gaza	85.3	1.06	0.012	83.2	87.4	3.6	6,029
Maputo Província	81.3	0.79	0.010	79.8	82.9	2.3	7,561
Maputo Cidade	72.5	1.14	0.016	70.3	74.7	3.0	8,411