

## Annex B: Survey methodology and data quality

The ultimate goal of the labour force survey (LFS) programme is to collect data on the employment and labour underutilization characteristics of the population on a continuous basis, providing quarterly estimates of the main labour force aggregates with sufficient precision at the district level. Given this ambitious measurement objective, it has been decided to adopt a gradual approach starting with an annual survey with the sample spread over four points in time. The sample is designed such that the survey provides for:

- Quarter national estimates of employment and labour underutilization with specified precision, and
- Annual district-level estimates by pooling together the data of each quarter of the calendar year.

The timing of the quarter rounds is based on the seasonal variations of labour force activities in Rwanda, in particular, the high and low seasons of agriculture activities, in February May , August, and November respectively. This approach is in line with the international standards that recommend a national data collection strategy that allows reporting “on a quarterly basis, main aggregates of employment, labour force, labour underutilization, including unemployment, in order to monitor short-term trends and seasonal variations.”

An annual survey with the sample spread over four points in time has a number of advantages in comparison with an approach that starts with an annual survey conducted at one point in time during the year. The quarterly approach allows NISR to put in place a survey management system that oversees survey operations and data processing tasks that cover the entire year. This would make it easier to transit at a later stage into a more frequent reporting system. It will also lessen the number of transitions and inevitable disruptions at each transition toward the final objective of reaching a continuous survey with quarterly reporting.

A quarterly approach also provides the possibility of incorporating a rotation scheme into the sample design. A rotation scheme allows to produce more accurate measurement of employment and labour underutilization trends as well as new statistics on labour force dynamics, in particular, flow data on job gains (number of persons obtaining employment during a given period) and corresponding data on job losses (number of persons losing employment during the period).

### 1. Sample design

The sample design of the LFS is a two-stage stratified design according to which at the first stage of sampling, a stratified sample of enumeration areas from the latest population census is drawn with probabilities proportional to size measured in terms of the census number of households or census number of household members, and at the second stage of sampling, a fixed number of sample of households is selected with equal probability within each sample enumeration areas. Finally, all household members in the sample households are selected for survey interviewing.

The scope of the survey is all persons living in private households. It excludes the institutional population permanently residing in houses such as hostels; health resorts; correctional

establishments etc., as well as persons living in seasonal dwellings not covered in the survey. It also excludes workers living at their work-sites. A household is a group of persons who live together and make common provision for food and other essentials for living. The people in the group may be related or unrelated or a combination of both. A household may consist of only one person or several persons.

- **Sample size**

Sample size determination in most household-based surveys with multi-stage stratified design is based on the principle of first calculating the required sample size for a single «domain» assuming a simple random sample design and no non-response. A domain is a well-defined population group for which estimates with pre-determined accuracy are sought. The results are then extended to allow for non-response and deviation from simple random sampling.

The application of this principle with the choice of parameters described below leads to a sample size requirement of 18,691 households for measuring annual unemployment with margin of errors of +/- 0.3% at 95% confidence level. In these calculations, the main indicator is the ratio of unemployment to the working age population, set at  $r=0.024$  according to the 2012 population census of Rwanda. The design effect (deff) is set at 3, within the range of values (3 to 4) generally used for labour force surveys. The margin of errors (ME = 0.0026) is twice the standard error of the estimate.<sup>17</sup> Similar calculations for employment gave a sample size requirement of 18,792 households.

The Labour Force survey programme begun with a pilot survey conducted in February 2016. The first round of the survey was conducted in August 2016, and continued on bi-annual basis until August 2018. Since the year 2019, the survey was re-designed to provide estimates of the labour market aggregates on quarterly basis. The sample size in each quarter is about 4608 households, composed of three rotation groups marked with the symbol  $r$  in the table. The proposed rotation design is 1-1-1, that is each sample household is interviewed three times, once every two quarters. Accordingly, a sample household is in the sample in one quarter, leaves the sample in the next quarter, returns in the sample in the following quarter, leaves again the sample in the subsequent quarter, and finally returns in the sample for the third time before leaving the sample altogether.

- **Sample rotation**

The main purpose of sample rotation is to improve the precision of the trend estimates. It also allows obtaining data on labour force flows by matching sample individuals common in different survey rounds. The sample rotation scheme adopted for the Rwanda LFS programme builds on the possibility of transiting to quarterly sub-rounds after a few years of half-annual sub-rounds.

---

<sup>17</sup>For more detail description of the calculations reference is made to: “Labour Force Survey of Rwanda: Proposed survey programme and instruments, FarhadMehran, GIZ Consultant, National Institute of Statistics of Rwanda (NISR), 31 July 2015.

## Rwanda LFS: Sample rotation design

Feb '19	May '19	Aug '19	Nov '19	Feb '20	May '20	Aug '20	Nov '20	Feb '21	May '21	Aug '21	Nov '21	Feb '22	May '22	Aug '22	Nov '22
r1															
	r2														
r3		r3													
	r4		r4												
(8+9)/3		r5		r5	r5										
	(8+9)/3		r6			r6									
		(10+11)/3		r7	r7		r7								
			(10+11)/3			r8		r8							
				(10+11)/3	(10+11)/3		r9		r9						
						(12+13)/3		r10		r10					
							r11		r11		r11				
								r12		r12		r12			
									r13		r13		r13		
										r14		r14		r14	
											r15		r15		r15
												r16		r16	
													r17		r17
														r18	
															r19

Source: National Institute of Statistics of Rwanda (NISR), Labour Force Survey

An implication of these modifications is a slight reduction of the sample size in August 2018 to ensure smooth transition to the sample size requirement of the quarterly design which must be multiple of 48 ( $=3 \times 16$ ), due to the transition from four to three visits per household and the maintenance of the sample-take of 16 households per PSU. The sample of the August 2018 survey has been supplemented by considering a sample-take of 24 households per PSU in one-third of the rotation groups 8 and 9 of the previous bi-annual design.

- **Sampling frame and sample allocation among districts**

A primary sampling unit is an enumeration area of the Population and Housing Census 2012. There are altogether 14,784 enumeration areas in the sampling frame constructed by NISR based on the population and housing census.

To ensure adequate geographical distribution of the sample over the different parts of the population, the sample is allocated among the 30 districts of the country. Prior to sample selection, the sampling frame is sorted according to urban and rural areas within districts. This provides an implicit stratification of the population by urban and rural areas. The resulting distribution of the sample is obtained on the basis of the square-root allocation.

It should be mentioned that the square-root allocation rule allocates the sample in proportion to the square-root value of the size of the district. It is a compromise between equal and proportional allocation. Equal allocation gives the same allocation to each district regardless of size. It may be appropriate if each district is to be treated as separate reporting unit (domain) with the same precision requirement as the others. Proportional allocation distributes the sample in proportion to the size of the districts. With proportional allocation, the geographical composition of the population is preserved, but it may lead to very small sample sizes for certain districts.

### **Selection of sample enumeration areas**

The next step in sample design was the selection of the 288 sample enumeration areas in each round. The sample was drawn in each district by probability proportional to size (pps) from the sampling frame.

The sample selection of PSUs has been carried out on the basis of the PSUs already selected under the bi-annual survey design. Thus, a sample of 144 PSUs have been selected from the 146 PSUs already designated for the August 2018 survey and what would have been the bi-annual February 2019 survey (old rotation groups 8 and 9). These PSUs have been randomly divided into three equal parts of each 46 PSUs, to serve for the new sample PSUs for introduction in the three consecutive quarters, August 2018, February 2019, and May 2019. Similarly, another sample of 144 PSUs have been selected from the 146 PSUs already designated for what would have been the bi-annual August 2019 and February 2020 survey (old rotation groups 10 and 11). These PSUs have then been randomly divided into three equal parts of each 46 PSUs, to serve for the new sample PSUs for introduction in the three consecutive quarters, August 2019, November 2019 and February 2020.

- **Selection sample households**

The sample enumeration areas were freshly listed prior to selection of the final sample of households. For each quarter, 16 sample households were selected from the list of households in each sample enumeration area by systematic sampling with equal probability. If the list contained 16 households or less all households in the sample enumeration area were drawn in the sample. If the list contained more than 16 households, a sample of fixed size (16 households) was drawn from the list by systematic random sampling. In order to keep almost the same number of households, the sample size in each quarter was about 4,668 households, composed of three rotation groups marked with the symbol r in the table1.

The purpose of listing was to ensure that every household currently residing in the sample locality has a non-zero probability of selection. Listing permits to update the sampling frame and account population movements and new household formations that have occurred since the last preparation of the sampling frame. It aims at covering the newly constructed buildings with living quarters and taking into account demolished or vacant buildings, or transformed dwellings no longer used as living quarters, such as dwelling addresses turned to stores or workshops, or living quarters used as secondary housing units or for holidays.

The listing form was developed to capture the information categorized in four columns: (1) A three-digit serial number for listing buildings; (2) Another three-digit serial number for listing dwellings within buildings; (3) The name of the head of household residing in the dwelling; and (4) The street address of the household.

- **Sample weights**

Three steps were involved in the calculation of the sample weights: Calculation of the design weight, Adjustment for non-response; and Calibration to known population projections.

The design weight of a given sample household is the inverse of the probability of selection of a sample household. This probability is calculated as the product of two probabilities. The first is the probability of selection of the enumeration area  $k$  where the sample household is residing,

$$p_k = \eta \times N_k$$

Where  $N_k$  is the number of households in the enumeration area according to the sampling frame and  $\eta$  is the proportionality factor of the pps sampling scheme.

The second is the probability of selection of the sample household within the enumeration area  $k$

$$p(hh_k) = \frac{16}{N_k^1}$$

where 16 is the fixed sample-take in enumeration area  $k$  and  $N_k^1$  is the listed or estimated number of households in the enumeration area  $k$ .

The overall design weight is the inverse of the product of these two probabilities,

$$DesignWeight(hh_k) = d_k = \frac{1}{p_k} \times \frac{1}{p(hh_k)}$$

The non-response adjusted weight is then obtained by the dividing the design weight with the response rate,

$$AdjustedWeight(hh_k) = d'_k = \frac{d_k}{r_k}$$

where the response rate  $r_k$  in enumeration area  $k$  is the percentage number of responding households among the total eligible households in the sample enumeration area.

The sampling weights for the quarterly LFS are calculated as a function of the weights calculated earlier for the bi-annual survey, called here "Oldweight" as described above. The Qweight is obtained from the Oldweight as follows,

$$Qweight = Oldweight \times \frac{3}{4} \times \frac{1}{144} \times \frac{1}{3} = Oldweight \times \frac{219}{96}$$

where  $\frac{3}{4}$  takes account of the fact that under the old bi-annual design there were 4 rotation groups per round and under the new quarterly design there are 3 rotation groups per round; the ratio 144/146 corresponds to the probability of selection of the new PSUs from the set of old PSUs; and, finally, 1/3 refers to the probability of selection when the new PSUs are randomly divided into three equal parts.

Finally, the adjusted weights were calibrated to known population projections for four demographic groups: Males and females population less than 16 years old and males and females 16 years old and over living on private households

The population projections were derived from the NISR census publication.<sup>18</sup> The projections were adjusted by deducting estimated values for the institutional population not living in private households. The calibration procedure followed the methodology of Deville and Sarndäl.<sup>19</sup> Accordingly, the final calibrated weights were obtained from the formula,

$$CalibratedWeight(hh_k) = w_k = d'_k \times (1 + \lambda x'_k)$$

where  $d'_k$  is the adjusted weight for non-response,  $\lambda$  is a regression vector obtained from the calibration formula, and  $x'_k$  is the vector of the count of male less than 16 years old, male 16 years old and over, female less than 16 years old and female 16 years old and over of interviewed households in the enumeration area k. All individuals in the same household are assigned the weight of the household in which they belong.

## 2. Questionnaire design

The questionnaire of the Rwanda Labour Force Survey 2018 in its present form contains a total of 149 questions organized into 9 sections and a cover page, dealing with following topics:

- A Household roster including activities of children aged 5 to 13 years old
- B Education
- C Identification of employed, time-related underemployed, unemployed and potential labour force
- D Characteristics of main job/activity
- E Characteristics of secondary job/activity
- F Past employment

<sup>18</sup> National Institute of Statistics of Rwanda, *Fourth Population and Housing Census, Rwanda, 2012, Thematic Report Population Projections*, January 2014.

<sup>19</sup> Deville, J.C., and Sarndäl, C.E., "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, Vol. 87, 1992, pp. 376-382.

- G Own-use production of goods and services
- H Subsistence foodstuff production
- I Housing and household assets

Not all questions are addressed to every household member. For children below 14 years of age, a minimum number of questions are asked. For older youngsters and adults 14 years of age and above, the number of questions depends on the situation and activities of the person during the reference period. The basic reference period is the last 7 days prior to the date of the interview. For certain questions, however, other reference periods are used. In each case, the relevant reference period is indicated in the text of the question.

The questionnaire was prepared both in Kinyarwanda and in English. An earlier version of the Kinyarwanda questionnaire was tested during the Pilot LFS February 2016. The field test was conducted in selected urban and rural areas with the aim of assessing the integrity of the instrument, such as understanding of question wordings, duration of interviews, coding and data processing. The experience gained was used to finalize the questionnaire.

Experience gained from the pilot survey led to certain modifications of the questionnaire. The revised questionnaire was again tested prior to the LFS August 2016 and February 2017 through the mock interviews conducted during the training of supervisors and interviewers. Slight modifications were further introduced following the experience gained during the field work of previous LFS rounds.

A specimen of the final version of the questionnaire is presented in Annex C of the present report. It is accompanied with two documents: An extensive manual for interviewers, providing instructions on the role of interviewers, listing of household members, and procedures to be adopted for asking each question and recording the corresponding response;<sup>20</sup> and a set of diagrams and corresponding STATA syntax providing rules for combining the survey responses for constructing the main labour force indicators of the survey. These derived variables included:

Labour force status (STATUS1):

Employed, Unemployed and Outside the labour force

Time-related underemployed (TRU)

Potential labour force (PLF)

Discouraged jobseeker (discourage)

Willing non-jobseeker (willing)

Subsistence foodstuff producer (sub)

Employment in informal and formal sector (IS and FS)

Informal and formal employment (IE and FE)

---

<sup>20</sup> NISR, *Rwanda Labour Force Survey, Interviewers Manual (RLFS, 2016)*, National Institute of Statistics of Rwanda, Kigali, December 2015. Revised for the RLFS, August 2016.

Monthly cash income from employment of employees at main job (cash)

Not in employment, nor in education or training youth 16-24 years old (NEET youth)

Not in employment, nor in education or training young persons 16-30 years old (NEET young)

Migrant worker (migrant)

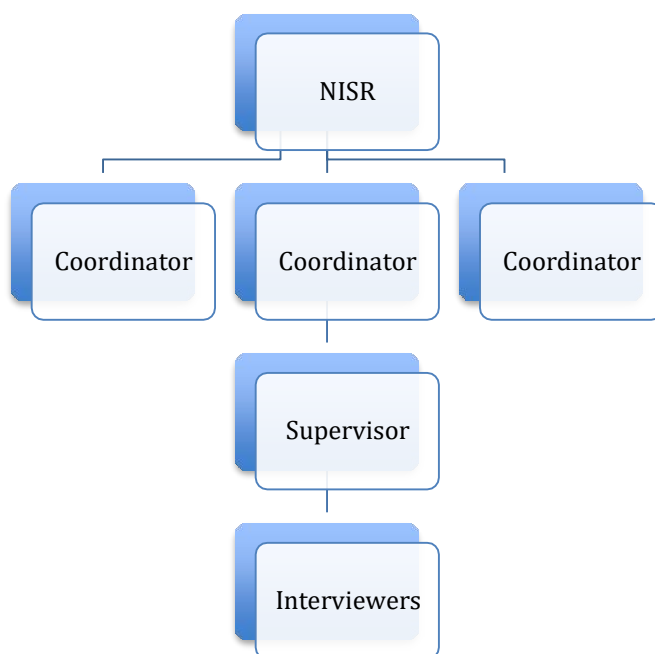
Worker with disability (disable)

### 3. Field operations

The main pre-survey activities conducted in preparation for the field operations included the establishment of the field organization, the recruitment and training of interviewers and the preparation of Tablets.

The following diagram shows the field organization of the survey. It consisted of 2 coordinators, coordinating the work of 10 supervisors, 24 team leaders and 96 interviewers. In many occasions the team leaders and field editors conducted also survey interviews. So altogether the survey interviewing was carried out by a total of 120 interviewers.

#### Field organization, LFS



The actual number of interviewers recruited for the survey was in excess of the required number in order to allow for drop-outs, illness, absences and other possible failures.

The training of the survey staff was carried out in two stages. There was first the training of the supervisors followed with the training of the interviewers. The coordinators conducted the training of the supervisors, and the combined pool of supervisors and coordinators conducted the training of the interviewers.



The field operations were conducted from February , May, August and November, each month from the beginning of the month up to the end of the month during the year 2019. Listing was conducted at the same time as data collection in Kigali city and others District out of Kigali.

Prior to household selection and interviewing, all households in the sample enumeration areas were listed in a special listing form. A separate manual was prepared for household listing, providing special instructions for segmentation of large clusters and quality control of the results.<sup>21</sup> Finally, procedures were developed for selecting sample households from the list as part of the field operations.<sup>22</sup>

The new 9 sample enumeration areas with very large number of households at the listing stage were segmented according to the segment areas defined by the population and housing census. One segment at random was listed and the number of listed households was proportionally expanded to obtain an estimate of the current number of households residing in the enumeration area.

#### **4. Data processing**

As well as from the previous year, data were collected using computerized assisted interview (CAPI). Data was uploaded to NISR main office from field via wireless network channel by synchronizing every day with the NISR server. It was carried every day to have a daily back up of data. All the activity of codification were also done to the field by interviewers who were trained. Several questions with textual responses were pre-coded and tabled in cascaded way. These concerned education (major field of study in highest qualification attained, and subject of training), occupation and branch of economic activity (at main and secondary job and past employment experience); they were coded into the corresponding national standard classifications using on-screen coding with corresponding dictionaries in Kinyarwanda. <sup>23</sup> Coding of geographic areas and addresses was incorporated in the data entry programme as look-up.

Following coding, responses of each questionnaire were edited for blanks, missing values, duplicates, out-of-range values, and inconsistencies such as no head of household or age of child greater than age of head of household using developed batches of controlling inconsistency in CsPro and Stata. Editing specifications on coverage and demographic characteristics were based on the population and housing census (PHC4 2012). Other edit rules were developed for consistency checks on questions related to the measurement of the main labour force variables, including employment, unemployment, multiple jobholding, total hours usually worked at all jobs, total hours actually

---

<sup>21</sup> National Institute of Statistics of Rwanda, *Manual for Household Listing, Rwanda Pilot Labour Force Survey (RLFS-P 2016)*, NISR, Kigali, February 2016.

<sup>22</sup>Mehran, F., GIZ Consultant, "Rwanda Labour Force Survey February 2016. Selection of households without data entry as part of the field operations." 30 December 2015.

<sup>23</sup> National Institute of Statistics of Rwanda, *Customized International Standard Industrial Classification of all Economic Activities (ISIC Rev. 4)*, The Rwanda Classification Manual, 2012 edition.

National Institute of Statistics of Rwanda, *Customized International Standard Classification for Occupation (ISCO-08)*, The Rwanda Classification Manual, 2012 edition.

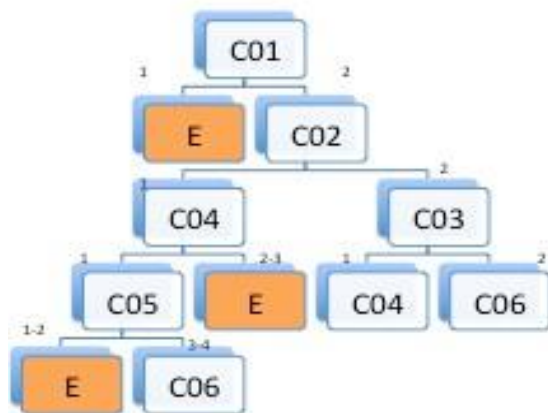
National Institute of Statistics of Rwanda, *Customized International Standard Classification of Education (ISCED 97)*, The Rwanda Classification Manual, 2012 edition.

worked at all jobs, status in employment at main job, etc. The detected errors were directly sent back to the field for the correction by the interview.

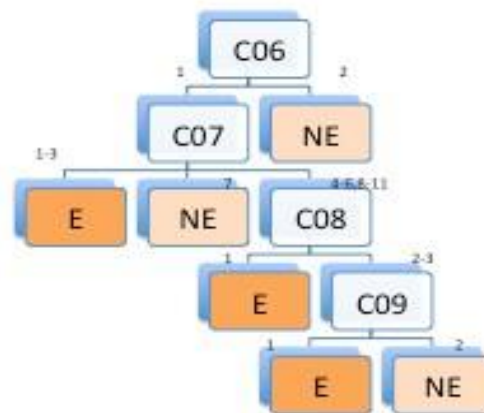
As part of data processing, the data file was augmented by adding a field on sampling weights (weight) and a series of additional fields on derived variables constructed on the basis of the information on each record. Some examples of the construction of the derived variables is schematically shown in diagrams B1 to B5 for employment (E), time-related underemployment (TRU), unemployment (U) and potential labour force (PLF), employment in informal sector (IS) and informal employment (IE) at main and secondary jobs, and monthly cash income from employment of employees at main job (cash). The numbered elements of the diagrams refer to the question numbers and response categories of the LFS questionnaire. The end nodes of the diagrams refer to the derived variable categories, employed, time-related underemployed, unemployed, etc.

## B.1 Derived variable: Employment (E)

### Employed, at work

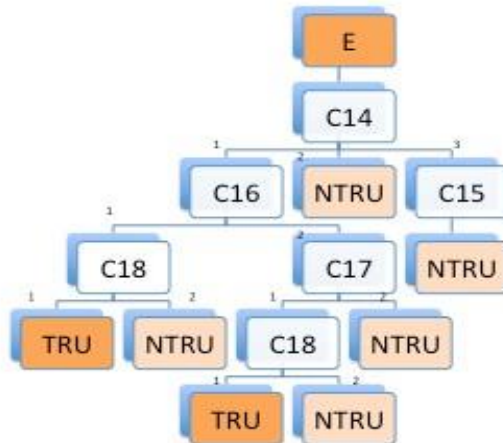


### Employed, not at work



E = Employed  
NE = Not employed

## B.2 Derived variable: Time-related underemployment (TRU)

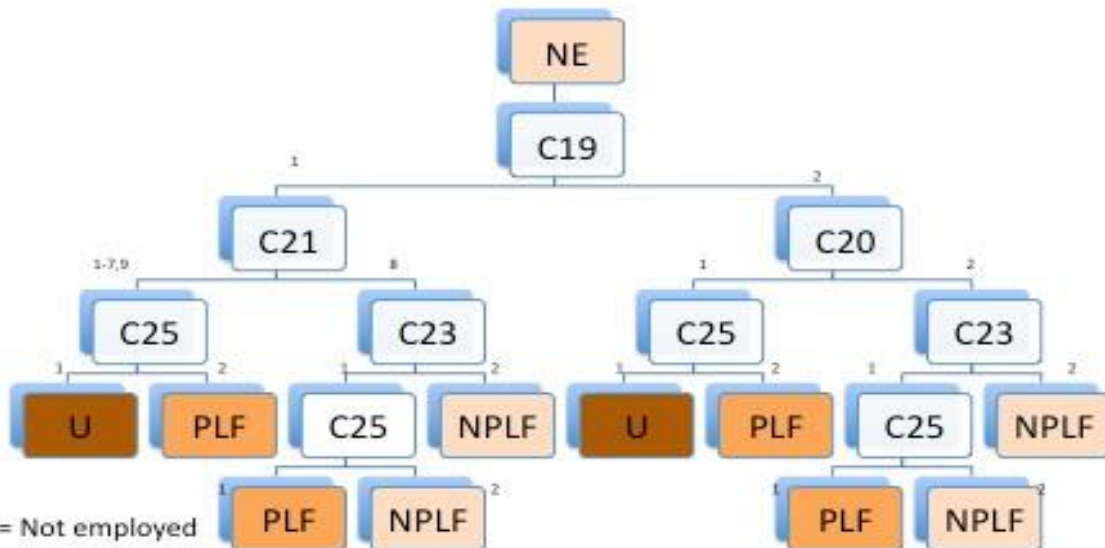


E = Employed

TRU = Time-related underemployed

NTRU = Not time-related underemployed

## B.3 Derived variables: Unemployment (U) and potential labour force (PLF)



NE = Not employed

U = Unemployed

PLF = Potential labour force

NPLF = Not in potential labour force

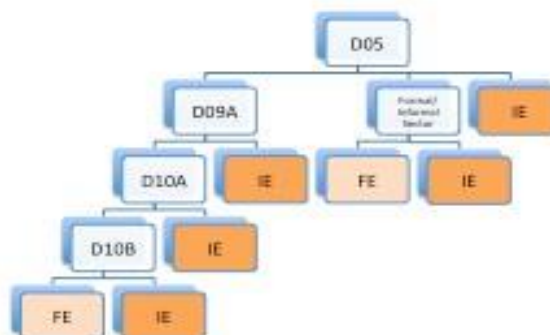
## B.4 Derived variables: Informal sector (IS) and informal employment (IE) at main job

### Informal sector



IS = Informal sector  
FS = Formal sector  
X = Out-of-scope (Domestic workers engaged by households are excluded from the count of informal and formal sector)

### Informal employment



IE = Informal employment  
FE = Formal employment

B.5 Derived variable: Monthly cash income from employment of employees at main job (INC)

Status in employment		Response	Monthly cash income from employment at main job
-1	-2	-3	-4
Employee or Intern or Other	D05=1,2,7	D12=1 Amount	D12A x 1 if D13=1
			D12A x 26/12 if D13=2
			D12A x 52/12 if D13=3
			D12A x 52/2 if D13=4
			D12A /12 if D13=5
		D12=2,3, Refusal, Don't know	2*20,000/3=13,333 if D17=1
			2/(1/20,000+1/30,000)=24,000 if D17=2
			2/(1/30,000 +1/50,000)=37,500 if D17=3
			2/(1/50,000 +1/100,000)=66,667 if D17=4
			2*100,000 = 200,000 if D17=5

Based on these results, it was decided to use the conversion factor 1 for monthly payments, 2 for two-weekly payments, 52/12 for weekly payments, and 26 for daily payments to calculate the monthly income.

Finally, the augmented data file with derived variables and sampling weights was used for producing the survey estimates specified in the tabulation programme of the survey as well as other analytical tables for the body of the report. The tabulation programme of the survey included 66 tables presented in the statistical annex of the present report (Annex D).

In order to speed up data processing and ensure better quality data, NISR has introduced tablets for data collection, and incorporated an assisted coding of the questions with textual responses and automated editing procedures for both detection and correction of errors thus minimizing the need for the time-consuming task of making reference to the physical questionnaires.

## 5. Data quality

Like in all sample surveys, the results of the LFS 2021 are subject to sampling and different forms of measurement errors. This section provides information on different sources of survey errors, namely, sampling errors, coverage errors, non-response errors, response errors and other errors such as coding and data entry errors.

### - Sampling errors

Sampling errors arise due to the fact that the survey did not cover all elements of the population, but only a selected portion. The sampling error of an estimate is calculated on the basis of the difference between the estimate and the value that would have been obtained on the basis of a complete count of the population under otherwise identical conditions.

Information on sampling errors is used for interpreting the survey results. It provides an assessment of the precision of the estimates and on the degree of confidence that may be attached to them. In the

same vein, it allows decision on the degree of detail with which the survey data may be meaningfully tabulated and analyzed. Information on sampling errors is also used for determining whether the survey estimates of change over time or the estimates of differences between two or more population subgroups are statistically significant. Finally, information on sampling errors may be used for future sample design. Rational decisions on the choice of sample size, sample allocation among strata, clustering and estimation procedures, can only be made on the basis of detail knowledge of their effect on the magnitude of sampling errors in the resulting statistics obtained from the survey.

The following table gives the sampling errors of the main labour force estimates obtained from the LFS 2021 . They have calculated based on the general principle that in multi-stage sample designs the variance contributed by the later stages of sampling is, under broad conditions, reflected in the observed variation among the sample results for first-stage units. Thus, the sampling variance of a variety of statistics, such as totals, means, ratios, proportions, and their differences can be obtained on the basis of totals calculated for the primary sampling units, here the localities 24. The calculations took into account the fact that the sampling weights were calibrated and used the residual method proposed by Deville and Sarndäl p. 380. They have been calculated using STATA syntaxes, taking into account of the survey design characteristics.

B.6: Sampling errors of estimates of main labour force aggregates

Indicator	Estimate	Standard error	Relative standard error	Confidence interval	
				Lower	Upper
Population 16+ yrs	7,718,871	114,600	1.5%	7,494,016	7,943,726
Labour force	4,166,625	69,989	1.7%	4,029,300	4,303,951
Employment	3,287,697	60,089	1.8%	3,169,797	3,405,596
Unemployment	878,929	28,325	3.2%	823,353	934,505
Outside labour force	3,552,246	67,424	1.9%	3,419,953	3,684,538

Source: National Institute of Statistics of Rwanda (NISR), Labour Force Survey, 2021

As an illustration of the use of the table, consider the second row of the table on the labour force. The total labour force, 4,166,625 is estimated with a relative standard error of 1.7 percent. The true value at 95 percent confidence level lies within the interval 4,029,300 and 4,303,951. Similarly, it can be stated that the total number of unemployed persons, 878,929 is estimated with a relative standard error of 3.2 percent. And, the true value at 95 percent confidence level lies within the interval 823,353 and 934,505.

The next table gives the estimated sampling errors for the main labour force indicators expressed in rates or percentages. For example the results indicate that the unemployment rate estimated at 21.1 percent has a standard error of 0.6 percentage points. This may be interpreted to mean that the true unemployment rate lies with 95 percent confidence within the interval, 20.0 and 22.2 percent.

<sup>24</sup>Verma, Vijay, *Sampling Methods*, Manual for Statistical Trainers Number 2, Statistical Institute for Asia and the Pacific (SIAP), Tokyo, Revised 2002.

B.7: Sampling errors of estimates of main labour force indicators (percentage)

Indicator	Estimate	Standard error	Confidence interval	
			Lower	Upper
Labour Force Participation Rate	54.0	0.5	53.0	54.9
Employment-Population Ratio	42.6	0.5	41.6	43.6
Unemployment rate	21.1	0.6	20.0	22.2

Source: National Institute of Statistics of Rwanda (NISR), Labour Force Survey, 2021

It is not practical to compute and report sampling errors for every published statistics of a labour force survey. For this purpose, general variance estimates are typically calculated using the approximate relationship between the variance of an estimate and its size, expressed by  $\text{var}(y)/y^2 = b + a/y$ . The results are presented in the following table.

B.8: Approximate sampling errors by size estimates

Size of estimate	Standard error	Relative standard error	Confidence interval	
			Lower	Upper
5,000,000	63,000	1.3%	4874000	5126000
2,500,000	37,000	1.5%	2426000	2574000
1,000,000	20,000	2.0%	960000	1040000
500,000	14,000	2.8%	472000	528000
250,000	9,000	3.6%	232000	268000
100,000	6,000	6.0%	88000	112000
50,000	4,000	8.0%	42000	58000
25,000	3,000	12.0%	19000	31000
10,000	2,000	20.0%	6000	14000

Source: National Institute of Statistics of Rwanda (NISR), Labour Force Survey, 2021

Thus, an estimate about 5,000,000 has an approximate standard error of 63,000 with a confidence interval at 95% level between 4,874,000 and 5,126,000. Similarly, an estimate of about 500,000 has an approximate standard error of 14,000 with a confidence interval between 472,000 and 528,000. It can be observed that the relative standard error sharply increases as the size of the estimate decreases. Estimates as low as 10,000 have very high relative standard errors, almost 20 percent. The table can be used to decide on the size of estimates that can be meaningfully considered or more as statistical significant for analysis. For size of estimates that are not listed in the table, the approximate standard errors can be obtained by interpolation or extrapolation of the values given in the table.

- **Non-response errors**

Non-response occurs due to failure to obtain the required information from the units selected in the sample (unit non-response) or to failure to obtain some items of information for the selected unit (item non-response). Unit non-response may occur due to incorrect address of the sample household, or inaccessibility of certain dwellings or refusal of the sample household to be interviewed, or because no one was at home when the interviewer contacted the household, or for other reasons.

Absence and refusal are considered as non-response while vacant demolished or out-of-scope housing units are considered as non-coverage. The non-response rates for all February, May, August and November round 2021 have been less than 9.4 percent.

- **Response errors**

Response errors refer to errors originating at the data collection stage. In relation to an individual respondent, response errors may occur because the respondent was unwilling to divulge certain information or because the respondent did not know the answer to the question asked or did not fully understand the meaning of the question. Response errors can also occur due to memory lapses, for example by forgetting to report an event, or incorrectly reporting the timing. Response errors may also occur because of errors made by the interviewer or by the instrument used for measurement. Interviewers may introduce errors because of haste and misreporting the responses, or because of misunderstanding of the survey concepts and procedures, or preconceptions and subjective biases. The questionnaire itself may be faulty, with wrong question wordings and incorrect skipping patterns.

The measurement of response errors is one of the most difficult parts of quality assessment of survey data. It generally requires carefully designed re-interview programmes. In the absence of such data, the quality of survey responses may be assessed by comparing the survey results with corresponding information from more reliable external sources such as administrative sources, for example, reconciling the LFS estimate of employment with the corresponding estimate obtained from the Integrated Business Enterprise Survey.<sup>25</sup> More detailed assessment may be carried out by comparing the LFS estimates of employment in specific occupations such as primary and secondary school teachers, nurses, and civil servants with corresponding statistics from the line ministries. Other indicators of response errors may be obtained by measuring the degree of self-response against proxy-response, or by testing the internal consistency of certain sets of inter-related responses. The assessment of response errors along the lines described here should be considered in future rounds of the survey when the survey programme has been stabilized.

- **Other errors**

Other sources of errors include coding and editing errors, as well as errors in data entry and data processing. Here the occupation and industry coding is evaluated. Table B.10 shows that the employed persons were coded in 359 distinct 4-digit occupation codes and 299 distinct 4-digit industry codes. The percentage of employed persons who were coded with codes ending with the

---

<sup>25</sup>National Institute of statistics of Rwanda, *Integrated Business Enterprise Survey*, NISR 2014.



digit “9” was 2.5 percent for occupation and 7.0 percent for branch of economic activity. Codes ending with “9” indicate that the occupation or industry descriptions in the LFS questionnaire could not be precisely found in the classification system and had to be coded as “other”, suggesting insufficient information for precise occupation and industry coding.

Table B.10: Coding into occupation and branch of economic activity codes ending with “9”

	<b>Occupation (ISCO-08)</b>	<b>Branch of economic activity (ISIC Rev 4)</b>
<b>February 2021</b>		
Distinct 4-digit codes	<b>252</b>	<b>220</b>
Number of employed persons coded in with 4-digit codes ending with “9” (non-weighted)	123	343
Total number of employed persons (non-weighted)	4,740	4,740
Percent	2.59%	7.24%
<b>May 2021</b>		
	Occupation (ISCO-08)	Branch of economic activity (ISIC Rev 4)
Distinct 4-digit codes	<b>259</b>	<b>206</b>
Number of employed persons coded in with 4-digit codes ending with “9” (non-weighted)	123	320
Total number of employed persons (non-weighted)	4,584	4,584
Percent	2.68%	6.98%
<b>August 2021</b>		
	Occupation (ISCO-08)	Branch of economic activity (ISIC Rev 4)
Distinct 4-digit codes	<b>255</b>	<b>220</b>
Number of employed persons coded in with 4-digit codes ending with “9” (non-weighted)	125	370
Total number of employed persons (non-weighted)	4,676	4,676
Percent	2.67%	7.91%
<b>November 2021</b>		
	Occupation (ISCO-08)	Branch of economic activity (ISIC Rev 4)
Distinct 4-digit codes	<b>238</b>	<b>207</b>
Number of employed persons coded in with 4-digit codes ending with “9” (non-weighted)	111	308
Total number of employed persons (non-weighted)	5,142	5,142
Percent	2.16%	5.99%

Source: National Institute of Statistics of Rwanda (NISR), Labour Force Survey, 2021

Overall, the quality assessment of the results of the LFS 2021 shows acceptable levels of sampling errors and non-response errors.

