Reference Manual

for Harmonizing Household Surveys

Survey-based Harmonized Indicator Program (SHIP)
Africa-PREM Statistics Practice Group

The World Bank
1818 H Street, N.W.
Washington, D.C.
20433, USA.

June 2013

Table of Contents

Chapter 0: Survey-based Harmonized Indicator Program (SHIP) 4

Chapter 0.1 SHIP Directory Organization and Naming Convention 5

Chapter 0.2 SHIP File Naming Convention 6

Chapter 0.3 SHIP Variable Naming Convention 7

Chapter 1: Module I – Harmonized Individual-Level Variables 7

Table 1.0 Sample, Geography and Basic Household Identifier (00_BASICINFO.dta) 7

Table 1.1 Basic Demographic Characteristics (CCD_YEAR_I.dta) 8

Table 1.2 Literacy and Education (CCD_YEAR_I.dta) 9

Table 1.3 Health (CCD_YEAR_I.dta) 11

Table 1.4 Child Survival and Anthropometry for Children Aged 0-60 Months (CCD_YEAR_I.dta) 12

Chapter 2: Module II – Harmonized Labor Force Variables 13

Table 2.1 Identification, Individual, Educational and Training Variables (CCD_YEAR_L.dta) 15

Table 2.2 Labor-force: vocational training (CCD_YEAR_L.dta) 17

Table 2.3 Screening questions and primary employment (CCD_YEAR_L.dta) 18

Table 2.4 Secondary Employment (CCD_YEAR_L.dta) 29

Chapter 3: Module III – harmonized Household-Level Variables 33

Table 1.0 Sample, Geography and Basic Household Identifier (CCD_YEAR_H.dta) 33

Table 3.1 Household Head Characteristics (CCD_YEAR_H.dta) 34

Table 3.2 Housing and Utilities (CCD_YEAR_H.dta) 35

Table 3.3 Ownership of Durable and Non-Durable Goods (CCD_YEAR_H.dta) 38

Table 3.4 Household Expenditure (CCD_YEAR_H.dta) 40

Table 3.5 Household Remittances (CCD_YEAR_H.dta) 41

Table 3.6 Household Income Variables (CCD_YEAR_H.dta) 42

Chapter 4: Module IV – Harmonized Expenditure Variables 43

Table 1.0: Household Composition and Weights Summary Table 44

Table 4.1: Food Expenditures 44

Table 4.2: Alcoholic Beverages, Tobacco and Narcotics 45

Table 4.3: Clothing and Footwear 45

Table 4.4: Housing, Water, Electricity, Gas and Other Fuels 45

Table 4.5: Furnishings, Household Equipment and Routine Household Maintenance 46

Table 4.6: Health Care Expenditures 46

Table 4.7: Transport Expenditures 47

Table 4.8: Communication Expenditures 47

Table 4.9: Recreation and Culture Expenditures 47

Table 4.10: Education Expenditures 48

Table 4.11: Restaurants and Hotels 48

Table 4.12: Miscellaneous Goods and Services 48

Table 4.13: Regional Price Deflators (OPTIONAL) 49

Table 4.14: Total Annual Household Expenditures (CCD_YEAR_E.dta) 50

Chapter 5: Documentation and Computer Programs 51

Appendix I: The Methodology of constructing employment variables in SHIP 52

Appendix II: Issues In Household Expenditure Aggregation 57

Data cleaning 57

Data Aggregation 57

Annex III: ISO 3166-1 Alpha-3 - Country Codes (Sub-Saharan Africa) 59

Annex IV: International Standard Industrial Classification (ISIC) 60

References 61

Appendix V: Standard Excel Spreadsheet 63

For any further inquiries please email xye@worldbank.org

CHAPTER 0: SURVEY-BASED HARMONIZED INDICATOR PROGRAM (SHIP)

As Sub-Saharan African economies become more open and globalized, huge opportunities are created for individuals and families. Yet, a large fraction of the population has not benefited sufficiently, and economic and social inequality is real and, in some cases, growing. Household surveys are a data source providing rich information on the impact of economic changes on individual households. This source of information, however, is largely underutilized due to the complexity of household surveys and significant time required to prepare the survey data for analytical work.

The SHIP seeks to eliminate the bottleneck of analyzing household survey data by extracting from existing household surveys about 200 SHIP variables that have the same definition and variable names. These variables include household consumption, access to infrastructure (water, electricity, education and health care), status of employment, education, and health.

This manual presents detailed guidelines for compiling household survey data into a set of most commonly available variables from various types of household surveys. To ensure the quality and transparency of the data, it is critical to document the procedures of compiling consumption aggregation and other indicators so that the results can be replicated from the original household survey data with ease. This enables data consistency and continuity that make temporal and cross-country comparisons consistent and more reliable.

Four harmonized data files are prepared for each survey to generate a set of harmonized variables that have the same variable names and definitions. Procedures used to compile these variables will be extensively documented. Invariably, in each survey, questions will be asked in different manner, which poses challenges to consistently define harmonized variables. The harmonized household survey data presents the best available variables with harmonized definitions. The four harmonized data files are:

a) **Module I: Individual file** (except Labor force variables): Information on basic characteristics of individuals such as age, sex, literacy, education, health, and child anthropometry.
b) **Module II: Labor force file**: Information on labor force including employment/unemployment, earnings, sectors of employment, etc.
c) **Module III: Household file**: Information on household expenditure, characteristics of household head (age, sex, level of education, employment), housing amenities, assets, and access to infrastructure and services.
d) **Module IV: Household Expenditure file:** Information on consumption/expenditure aggregates organized by consumption groups according to the UN Purpose of Household Consumption (COICOP).

Before harmonizing a household survey using this manual, the user should first create 6 sub-directories as instructed below. Additionally, all SHIP files must be named according to this manual. It is critical that all data editing checks (cleaning and assumptions) should be documented in detail in the computer programs as comments. This rigorous procedure is to ensure that the SHIP files are replicable from the original survey data, which proves to be invaluable when SHIP files are used for analytic work.

CHAPTER 0.1 SHIP DIRECTORY ORGANIZATION AND NAMING CONVENTION

Six sub-directories should be created under the main directory named after the survey, as shown below.

Main directory:	**Storage Purposes**
"\COUNTRYCODE_SURVEYNAME_YEAR\"	Parent directory of all sub-directories
Sub-Directories	
\01_RAWDATA	All survey data sets as keyed in (no editing has been done on this data). All questionnaires, survey manuals and any related documents should also be kept in this directory. This directory guards the original survey directory from the National Statistical Offices, and will never be touched during the SHIP process.
\02_CLEANDATA	All clean survey data sets before aggregation. SHIP extracts all its SHIP files from this directory.
\03_PROGRAMS	All program files that derive the SHIP indicators (See SHIP file naming convention below).
\04_AGGDATA	All SHIP output data files produced by the above programming files(see file naming convention below)
\04_AGGDATA_PUBLIC	All SHIP files with anonymized household ID
\05_DOC	All reports related to the survey found from other sources, such as poverty assessment, etc.

CHAPTER 0.2 SHIP FILE NAMING CONVENTION

For each survey, the SHIP creates 21 program files, 18 data files (and another set of 18 data files with scrambled household ID, but all other variables are the exactly the same). They should be all named according to the table below (CCD stands for ISO 3 letter country code, see Appendix III). The objective of the file naming convention is to enable the complete documentation and transparency of extracting SHIP variables from the original household survey data. All SHIP files should be replicable from the original data based on SHIP program files. The extension of the files for different computer languages will be different (In this manual, all file extensions are named using STATA language). The file names for the computer programs and their output data files should be the same, ensuring an easy tracking between the SHIP data files and their original household survey files.

SHIP Program files	SHIP output (data) files	Corresponding tables in the manual
00_CCD_BASICINFO.do	00_CCD_BASICINFO.dta	Table 1.0
01_CCD_EXPFOOD.do	01_CCD_EXPFOOD.dta	Table 4.1
02_CCD_EXPALCH.do	02_CCD_EXPALCH.dta	Table 4.2
03_CCD_EXPCLTH.do	03_CCD_EXPCLTH.dta	Table 4.3
04_CCD_EXPHOUS.do	04_CCD_EXPHOUS.dta	Table 4.4
05_CCD_EXPFURN.do	05_CCD_EXPFURN.dta	Table 4.5
06_CCD_EXPHLTH.do	06_CCD_EXPHLTH.dta	Table 4.6
07_CCD_EXPTRSP.do	07_CCD_EXPTRSP.dta	Table 4.7
08_CCD_EXPCMNQ.do	08_CCD_EXPCMNQ.dta	Table 4.8
09_CCD_EXPRCRE.do	09_CCD_EXPRCRE.dta	Table 4.9
10_CCD_EXPEDU.do	10_CCD_EXPEDU.dta	Table 4.10
11_CCD_EXPHOTL.do	11_CCD_EXPHOTL.dta	Table 4.11
12_CCD_EXPMISC.do	12_CCD_EXPMISC.dta	Table 4.12
13_CCD_FDPINDEX.do	13_CCD_FDPINDEX.dta	Tables 4.13.1 and 4.13.2. (Optional)
CCD_YEAR_E.do	CCD_YEAR_E.dta	Variables from all above SHIP files
CCD_YEAR_CLEANING.do		Cleaning programing for the CCD_YEAR_E.dta
CCD_YEAR_I.do	CCD_YEAR_I.dta	Tables 1.0, 1.1, 1.2, 1.3, and 1.4
CCD_YEAR_L.do	CCD_YEAR_L.dta	Tables 1.0, 2.1, 2.2, 2.3, and 2.4
CCD_YEAR_H.do	CCD_YEAR_H.dta	Tables 1.0, 3.1, 3.3, 3.4, 3.5, and 3.6
CCD_YEAR_INDICATORS.do	INDICATORS.dta	Extract 60 indicators from all SHIP files by consumption quintile and rural urban areas
Z_CCD_YEAR_ID.do	Z_CCD_YEAR_ID.dta and all of the above files with a suffix "_PUBLIC"	

CHAPTER 0.3 SHIP VARIABLE NAMING CONVENTION

All 200 plus SHIP variables are listed in the four modules below, along with their definitions and the way to compile them. Variables in the data files must follow the sequence in which they appear in the tables. Not all SHIP variables are available in all surveys. When there are no questions in the survey corresponding to a particular SHIP variable, the variable should still be created with missing values. However, if there is no labor force module in a survey, then the entire labor force file (CCD_YEAR_L.dta) should be missing.

CHAPTER 1: MODULE I – HARMONIZED INDIVIDUAL-LEVEL VARIABLES

This module extracts variables of individuals in the household and covers approximately forty quantitative variables. The information is organized in 5 tables that provide variables on basic household identification, demographic characteristics, education, health, and child's vaccination and anthropometry.

TABLE 1.0 SAMPLE, GEOGRAPHY AND BASIC HOUSEHOLD IDENTIFIER (00_BASICINFO.DTA)

These variables will be included in all SHIP files except for the variable DISTRICT

No	SHIP VARIABLE NAME	LABEL AND CODES	COMMENTS, QUALITY CONTROL CHECKS AND CORRECTIONS
1	**COUNTRY**	**Country code**	String variable, of 3 character length (Appendix II
2	**REGION**	**Region code and label the names**	Codes vary by country
2b	**PROVINCE**	**Province code and label the names**	Codes vary by country.
2c	**DISTRICT**	**District code and label the names (if available)**	Codes vary by country.
3	**HID**	**Household unique identification**	**String variable**, length varies by country.
4a	**SURVEMO**	**Month of survey**	The month when the questionnaire was administered.
4b	**SURVEYR**	**Year of survey**	The year when the questionnaire was administered.
5	**RURURB**	**Area of residence** *1 = Rural* *2 = Urban*	
6	**HHSIZE**	**Household members**	Total # of residents (regular members) in the household, including maids and servants
7	**WTA_S**	**Individual weighting coefficient**	Weights from sampling, used in weighting statistics calculated from individual level variables.
8	**WTA_S_HHSIZE**	**Household weighting coefficient**	WTA_S multiplied by HHSIZE, normally used in weighting statistics calculated from household level variables.

TABLE 1.1 BASIC DEMOGRAPHIC CHARACTERISTICS (CCD_YEAR_I.DTA)

No	SHIP VARIABLE NAME	LABEL AND CODES	COMMENTS, QUALITY CONTROL CHECKS AND CORRECTIONS
I1	**INDID**	**Individual identification**	A string variable uniquely identifies the household members in each household. Sequentially numbered from 1 for the household head to *N* (household size).
I2	**SEX**	**Sex** *1 = Male* *0 = Female*	Sex of the individual
I3	**RELAT**	**Relationship to household Head** *1 = Head* *2 = Spouse* *3 = Child* *4 = Father/Mother* *5 = Grandchild* *6 = Son-in-law/daughter-in-law* *7 = Other relative* *8 = Domestic help/paying boarder* *9 = None relative*	Relationship of each household member to the household head. Each household must have only ***one*** head.
I4	**HEAD**	**Household Head** 0 = No 1 = Yes	Assigned based on the reported household head status. Check to ascertain that a household has only **one** unique household head
I5	**MARSTAT**	**Marital status** *1 = Never married* *2 = Married monogamous* *3 = Married polygamous* *4 =Common law / union coutumiere / union libre / living together* *5 = Divorced / Separated* *6 = Widowed*	Marital status of individual. If polygamous status is not reported, check to find polygamous status in a household by counting the number of spouses in the household.
I6	**SP_PRES**	**Spouse present** *1 = Yes* *0 = No*	Code on the basis of a question that asks whether or not the spouse lives in the household. Some surveys, instead ask for length of absence. If absence is more than 6 months code as 0 (No).
I7	**AGEY**	**Age in completed years**	Ages **>=** 99 must be coded as 98. *Missing values* should be left as missing. For children **<=** 60 months, check consistency with age in months to ascertain correct age in completed years.

I8	**ETHNICITY**	**Ethnicity code and name**	Codes vary by country.
I9	**LANGUAGE**	**Language usually speak, code and label the names**	Codes vary by country.
I10	**RELIGION**	**Religion, code and label the names**	Codes vary by country.

TABLE 1.2 LITERACY AND EDUCATION (CCD_YEAR_I.DTA)

No	SHIP VARIABLE NAME	LABEL AND CODES	COMMENTS, QUALITY CONTROL CHECKS AND CORRECTIONS
I11	**LITERACY**	**Literacy status** *1 = Can read and write* *2 = Cannot read or write* *3 = Cannot be determined*	For individuals aged 5 and above only. Value must be missing for all others. Be careful while coding 1 (read **and** write)
I12	**EVERATTD**	**Ever attended school** *1 = Yes* *0 = No*	For individuals aged 5 and above only. Value must be missing for others.
I13	**ATSCHOOL**	**Currently enrolled in school** *1 = Yes* *0 = No*	Use the question that asks for current attendance. If such a question is missing, use the question that explicitly asks for enrollment over the past 12 months. In such surveys, record this in the comments. Code as 0 if EVERATTD is 0.
I14	**EDLEVEL_AR**	**Highest level of education completed** *0 = No education* *1 = Pre-school/ Primary not completed* *2 = Completed primary, but less than completed lower secondary* *3 = Completed lower secondary (or post-primary vocational education) but less than completed upper secondary* *4 = Completed upper secondary (or extended vocational/technical education)* *5 = Post secondary technical* *6 = University and higher* *7 = Formal adult education or literacy program* *9 = Other*	Code for all individuals who were administered in the Education section. If a person is currently enrolled in the highest year of education, then his/her level of education completed should be determined by minus one year. For example, if a person is currently enrolled in P6, then his/her highest level completed should be coded as 1 (Pre-school/ Primary, not completed). Individuals enrolled in higher/ University level are coded as 6 (University and higher) regardless of whether it is completed or not. If there are more than 1 percent of missing values please report to the supervisor for further instructions.

I15	**EDYEARS**	***Years of completed education*** 0 = pre-school 1 = grade 1 2 = grade 2 . . . 99 = unspecified/incomparable	Code for all individuals who were administered in the Education section. For individuals who are currently enrolled in school, their years of education completed correspond to the class *currently attending minus one*. For individuals who are not currently enrolled in school, the years of completed education corresponds to the highest level of education completed. This is a continuous variable of the number of years of formal schooling completed. It is constructed only if the survey asked for the number of year of education or highest grade level completed; otherwise, the values are constructed as missing. The years of education that each grade corresponds to, varies by country, for example - some countries may have 5 or 6 years of primary school, 3 years of lower-secondary school, while other countries may have 4 years of primary school and 4 years of lower-secondary school. Refer to the UNESCO ISCED mappings.[1] For higher education, the grades/years may not have been asked explicitly. In such cases, the variable should be constructed based on the following assumptions:- **1)** If the individual has completed the tertiary education specified, add to years of completed education - 4 years for BA/BSc, 6 years for MA/MSc, and 8 Years for PhD after the completion of secondary education. **2)** If the individual has not completed tertiary education or completion cannot be ascertained, add to years of completed education – 2 years for BA/BSc, 5 years for MA/MSc, and 7 years for PhD. The variable does not take into account the actual number of years required to reach this grade level. In other words, first grade repeated three times only counts as 1 year of completed education.

[1] http://www.uis.unesco.org/Education/ISCEDMappings/Pages/default.aspx

I16	**EDYEARS_AT**	Years of education for currently enrolled 0 = pre-school 1 = grade 1 2 = grade 2 . . . 99 = unspecified/incomparable	Code this variable only for individuals currently enrolled in school. Otherwise code as missing. Code the years of education corresponding to the currently attending grade.
I17	**SCHLTYP**	**Type of school attending** *1 = Public* *2 = Private* *9 = Other*	Code only for individuals currently attending school (ATSCHOOL=1). Value must be missing for others.

TABLE 1.3 HEALTH (CCD_YEAR_I.DTA)

No	SHIP VARIABLE NAME	LABEL and CODES	COMMENTS, QUALITY CONTROL CHECKS AND CORRECTIONS
I18	**MORB_TR**	**Time reference for variable MORBID (in weeks)**	Number of weeks. By convention, 1 month = 4 weeks
I19	**MORBID**	**Morbidity last MORB_TR** *1 = Yes* *0 = No*	Has the person been sick or injured in the past X (MORB_TR) weeks?
I20	**RTREATM**	**Sought medical attention** *1 = Yes* *0 = No*	Code only for sick individuals (MORBID=1).
I21	**HLTHFAC**	**Health provider visited** *1 = Hospital* *2 = Clinics* *3 = Dispensary, health center, health post* *4 = Outreach* *5 = Pharmacy* *6 = Traditional healer* *9 = Other*	Code for all individuals who reported it in the survey irrespective of whether they sought medical attention.
I22	**OWHPROVD**	**Ownership of health provider visited** *1 = Public* *2 = Private - modern medicine* *3 = Private - traditional healers* *4 = Missionary / NGO* *9 = Other*	Code only for all individuals who visited a health care facility (1<=HLTHFAC<=9).

TABLE 1.4 CHILD SURVIVAL AND ANTHROPOMETRY FOR CHILDREN AGED 0-60 MONTHS (CCD_YEAR_I.DTA)

No	SHIP VARIABLE NAME	LABEL AND CODES	COMMENTS, QUALITY CONTROL CHECKS AND CORRECTIONS
I23	**AGEM**	**Age of child in completed**	Only for children aged 0 - 60 months (in some surveys the cu-off months is 59 months); otherwise code as missing
I24	**WHDELIV**	**Who helped assist Childbirth** *1 = Trained Birth Attendant* *2 = Traditional Birth Attendant* *3 = Relative/friend* *9 = Other*	Only for children aged 0-60 months.
I25	**VCARD**	**Child has vaccine card** *1 = Yes* *2 = No, but child has been vaccinated* *3 = No, child never vaccinated*	
I26	**BCG**	**Child immunized against BCG** *1 = Yes* *0 = No* *9 = Don't know*	Only for children aged 0-60 months. BCG is the vaccine against tuberculosis.
I27	**POLIO_DO**	**Number of doses of Polio received** *0 = No dose* *1 = One dose* *2 = Two doses* *3 = Three doses* *4 = More than three doses* *9 = Unspecified number of doses*	Only for children aged 0-60 months. Vaccine is also referred to as OPV. If Polio vaccine has been given but the number of doses is unknown or is not specified, code as 9.
I28	**DPT_DOSE**	**Number of doses of DPT received** *0 = No dose* *1 = One dose* *2 = Two doses* *3 = Three doses* *4 = More than three doses* *9 = Unspecified number of doses*	Only for children aged 0-60 months. 3 doses refer to full immunization against Diphtheria, Pertussis and Tetanus.

I29	**MEASLES**	**Child immunized against measles** *1 = Yes* *0 = No* *9 = Don't know*	Only for children aged 0-60 months.
I30	**VACCINE**	**Immunization status of Child** *1 = Fully immunized* *2 = Partially immunized* *3 = Immunized but no information on completeness* *4 = Not immunized*	Only for children aged 0-60 months. *Fully immunized* refers to:- (a) BCG (1 or more doses). (b) Three or more doses of Polio (1-2-3). 3 or more doses may include Polio at birth (Polio-0). (c) Three or more doses of DPT (1-2-3) and (d) Measles (1 or more doses).
I31	**WEIGHT**	**Weight of Child (kg)**	Only for children aged 0-60 months.
I32	**HEIGHT**	**Height of Child (cms)**	Only for children aged 0-60 months. By convention 1 cm = 10 mms 1 inch = 2.54 cms.

Note:

a) For any variable not collected in a country, variable should be created and left as (.) in the final harmonized file.
b) Variables in the data files must follow the sequence in which they appear in the manual.

CHAPTER 2: MODULE II – HARMONIZED LABOR FORCE VARIABLES

Note: The construction of employment and labor participation variables is specific to Sub-Saharan African context since over 80 percent of employment activities are in the informal sector. Studying labor participation in Tanzania, Bardasi, Beegle, Dillon and Serneel found that due to poor questionnaire design many unpaid family workers under reported their economic activities, especially women who reported domestic duties as the main activity. These individuals inevitably undertake some unpaid economic activities such as cultivating, raising livestock, fetching water, and collecting wood (preparing meals for the family and caring for own children are not classified as economic activities by ILO definition. The SHIP developed complementary steps to capture these under-reported economic activities.

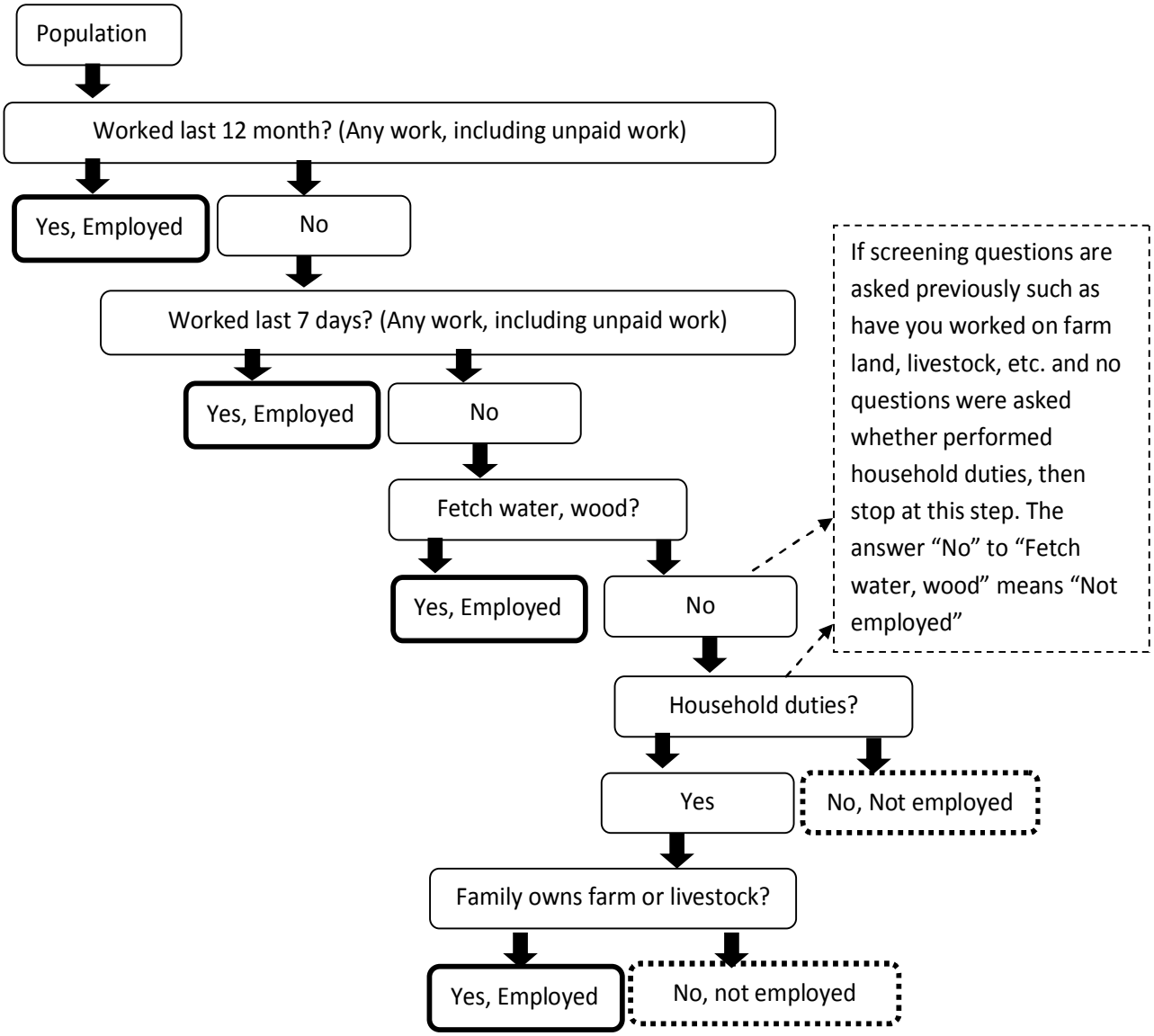
This manual reclassifies the employment status of these individuals, who claim household duties as their main activity, as employed rather than inactive. Because labor force questionnaires are significantly different from one another, it is not possible to provide a set of very specific steps that one can follow to classify employment status in Africa. The diagram below illustrates the logic used to classify unpaid economic activities. For details refer to Bardasi, Beegle, Dillon and Serneels "Do Labor Statistics Depend on How and to Whom the Questions Are Asked", World Bank Policy Research Working Paper 5192.

Additionally, because informal economic activities are so common, there are often missing values for the sectors of employment. To remedy this situation we designed a variable called "EMPFRM_1" and "EMPFRM_2" classifying primary and secondary employment respectively into two categories "Farm" and "Nonfarm" activities. Everyone who had a job in last year or last 7 days, must be classified into either Farm

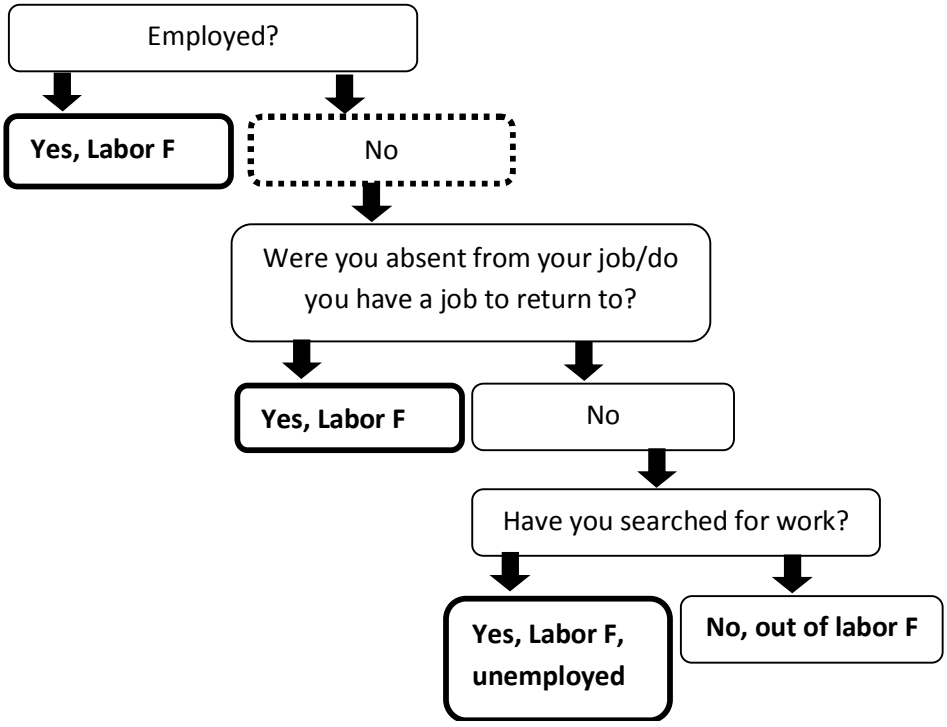
or Nonfarm activities using any information available in the survey that might help in this classification (less than one percent missing values is allowed).

It is also strongly recommended that the user studies Appendix I carefully before starting to construct labor force variables, for detailed and in-depth explanations of the logic and how to construct various key SHIP labor force variables.

Definition of Employment in Africa Region



Definition of Unemployment and Labor Force in Africa Region



(Variables below with sequence number starting with "L" are labor force variables; other variables are imported from other Individual/ household level SHIP files. Generate for all individuals)

All variables in this section to be included in CCD_YEAR_L.dta

TABLE 2.1 IDENTIFICATION, INDIVIDUAL, EDUCATIONAL AND TRAINING VARIABLES (CCD_YEAR_L.DTA)

No	NAME	LABEL AND CODES	COMMENTS, QUALITY CONTROL CHECKS AND CORRECTIONS
1	**COUNTRY**	**Country code**	To be merged from 00_BASICINFO.dta
2	**REGION**	**Region code**	To be merged from 00_BASICINFO.dta
2b	**PROVINCE**	**Province code**	To be merged from 00_BASICINFO.dta
2c	**DISTRICT**	**District code**	To be merged from 00_BASICINFO.dta
3	**HID**	**Household unique identification**	To be merged from 00_BASICINFO.dta
4a	**SURVEMO**	**Month of survey**	To be merged from 00_BASICINFO.dta
4b	**SURVEYR**	**Year of survey**	To be merged from 00_BASICINFO.dta
5	**RURURB**	**Area of residence** *1 = Rural* *2 = Urban*	To be merged from 00_BASICINFO.dta
7	**WTA_S**	**Individual weighting coefficient**	Weights from sampling. To be merged from 00_BASICINFO.dta.
I1	**INDID**	**Individual identification**	To be merged from CCD_YEAR_I.dta.

I2	**SEX**	**Sex** *1 = Male* *0 = Female*	To be merged from CCD_YEAR_I.dta.
I3	**RELAT**	**Relationship to household Head** *1 = Head* *2 = Spouse* *3 = Child* *4 = Father/Mother* *5 = Grandchild* *6 = Son/daughter-in-law* *7 = Other relative* *8 = Domestic help/paying boarder* *9 = None relative*	To be merged from CCD_YEAR_I.dta.
I4	**HEAD**	**Household Head** *0 = No* *1 = Yes*	To be merged from CCD_YEAR_I.dta.
I5	**MARSTAT**	**Marital status** *1 = Never married* *2 = Married monogamous* *3 = Married polygamous* *4 =Common law / union coutumière / union libre / living together* *5 = Divorced/ Separated* *6 = Widowed*	To be merged from CCD_YEAR_I.dta.
I7	**AGEY**	**Age in completed years**	To be merged from CCD_YEAR_I.dta.
I11	**LITERACY**	**Literacy status** *1 = Can read and write* *2 = Cannot read or write* *3 = Cannot be determined*	To be merged from CCD_YEAR_I.dta.
I12	**EVERATTD**	**Ever attended school** *0 = No* *1 = Yes*	To be merged from CCD_YEAR_I.dta.
I14	**EDLEVEL_AR**	**Education level** *0 = No education* *1 = Pre-school Primary, incomplete* *2 = Primary completed*	To be merged from CCD_YEAR_I.dta.

		but less than lower secondary completed *3 = Lower secondary completed (or post-primary vocational education) but less than upper secondary completed* *4 = Completed upper secondary* *5 = University and higher* *6 = Any post-secondary technical education* *7 = Adult education or literacy program* *9 = Other*	
I15	**EDYEARS**	**Years of education** *0 = pre-school* *1 = grade 1* *2 = grade 2.* . *99 = unspecified / incomparable*	To be merged from CCD_YEAR_I.dta.
E65	**FDPINDEX**	**Regional food price deflators**	To be merged from CCD_YEAR_E.dta
E70	**CPI2005_DEF**	**CPI Deflator**	To be merged from CCD_YEAR_E.dta
E71	**PPP2005**	**2005 PPP exchange rate**	To be merged from CCD_YEAR_E.dta

TABLE 2.2 LABOR-FORCE: VOCATIONAL TRAINING (CCD_YEAR_L.DTA)

No	NAME	LABEL AND CODES	COMMENTS, QUALITY CONTROL CHECKS AND CORRECTIONS
L1	**EMPLEDU**	**Post school vocational training** *0 = No* *1 = Yes*	Code as 1 (YES) if the individual has ever attended or is currently attending post school vocational training, otherwise 0 (NO).
L2	**SVYAGE**	**Lower age cut off for employment svy**	The lower age cut-off for individuals who were administered the LF questionnaire.
L3	**EMPLEDU_TU**	**Time Unit of course duration** *1=Days* *2=Weeks* *3=Months* *4=Years*	Categorical variables to capture the unit of time for the duration of the course.
L4	**EMPLEDU_TIME**	**Number of EMPLEDU_TU**	Continuous variable to capture the duration of the course.

L5	**EMPLEDU_CER**	**Highest certificate obtained**	The name of the certificate earned. Codes will vary with the country
L6	**EMPLEDU_SCHLTYP**	**Type of Training Institute** 1=*Public* *2=Private* *9=Other*	Code only if EMPLEDU=1. Leave as missing if EMPLEDU is not equal to 1.
L7	**APPREN**	**Apprenticeship** *0 = No* *1 = Yes*	Code as 1 (YES), if the individual **HAD** participated in an apprenticeship. *Note: If currently working as an apprentice, **DO NOT** code as 1 here (**code as 0**). It will be included in variable EMPTYPE_WB_1 and/or EMPTYPE_WB_2*

TABLE 2.3 SCREENING QUESTIONS AND PRIMARY EMPLOYMENT (CCD_YEAR_L.DTA)

No	NAME	LABEL AND CODES	COMMENTS, QUALITY CONTROL CHECKS AND CORRECTIONS
L8	**WORKED_Y**	**Labor force screening <u>last year</u>** *0 = No* *1 = Yes* *99=The question is not in the survey*	Use 12 month employment section only. Code as 1 (YES), if the individual performed any economic activity for cash or in-kind payment or barter or self-consumption for any length of time during the past 12 month, otherwise code as 0 (NO).
L9	**WORKED _7**	**Labor force screening <u>last 7 days</u>** *0 = No* *1 = Yes* *99= The question is not in the survey*	Use 7 day employment section only. Code as 1 (YES), if the individual performed an economic activity for cash or in-kind payment or barter or self-consumption for any length of time during last 7 days, otherwise code as 0 (NO). Code as 0 (No), if a person has a job but was absent for last 7 days or more. Do not re-code missing values.
L10	**ABSENT**	**Absent for 7 days or more** *0 = No* *1 = Yes* *99 = The question is not in the survey*	Use 7 day employment section only. **If WORKED_7 = 1 then code as missing value** (skip the variable) If WORKED_7 = 0, then code accordingly Code as 1 (YES), if the person has a job but was absent from it for any reason (sick, other leave, temp. layoff, is about to start work etc.), otherwise code as 0 (NO). Thus all individuals who did not work in the past 7 days **AND** do not have a job to go back to are coded as 0(NO).

L11	**LOOKJOB_TU**	**Up to xx weeks**	If both a 7 day and 12 month employment section is present in a survey, use only the 7 day employment section to code this variable Record the reference time period (in weeks) for which the question is asked. If the question gives a range of reference period, take the longest time period, up to 4 weeks. 4 weeks is the time period defined by the ILO *Note: (7 days = 1 week)*
L12	**LOOKJOB**	**Looked for job in last "LOOKJOB_TU" weeks** *0 = No* *1 = Yes* *99 = The question is not in the survey*	If both a 7 day and 12 month employment section is present in a survey, use only the 7 day employment section to code this variable. Code as 1 (YES) if **BOTH** the following conditions are satisfied, otherwise code as 0 (NO). 1. The individual did not perform an economic activity for cash or in-kind payment in the last 7 days. 2. The individual looked for a job in the LOOKJOB_TU weeks.
L13	**FETCHWOOD**	**Fetched wood for the household** *0= No* *1= Yes* *99 = The question is not in the survey*	Code as 1 (YES) if the individual fetched wood for his/her own household *or for others*, otherwise code 0 (NO). Based on UN definition of SSN, Fetching wood (*for pay or in-kind*) is an economic activity.
L14	**FETCHWATER**	**Fetched water for the household** *0= No* *1= Yes* *99 = The question is not in the survey*	Code as 1 (YES) if the individual fetched water for his/her own household *or for others*, otherwise code 0 (NO). Based on UN definition of SSN, Fetching water (*for pay or in-kind*) is an economic activity.
L15	**NOTLAB_WB**	**Reason not in labor force** *0 = In labor force* *1 = Household duties* *2 = Retired* *3 = Student* *4 = Dependent (too old*	If LABFORCE_WB=1 then code as 0, otherwise if LABFORCE_WB=0 then code according to responses. Code as 4 (Dependent), if the response is 'dependent' or 'too old' or 'too young' or if the individual is **below 5 years of age**.

		or too young to work) *5 = Disabled/ill* *6 = Discouraged worker* *9 = Other*	Use information from all sections (12 months and 7 days) that ask this question. First use information from the 12 month section. For all missing values and observations coded as 9, use the 7 day section to recode the variable.
L16	**EMPLHOME**	**Employment in household duties in own house** 0 = No 1 = Yes 99= The question is not in the survey	This variable is created to capture those individuals who were incorrectly classified as out of the labor force because in the questionnaire design, they had an option of reporting 'domestic duties' as their activity. Construct this variable in 2 stages: **Stage1: Identifying individuals engaged in household duties** An individual is classified as engaged in household duties if **ALL** the following conditions are met: 1. Did not work in the past 1 year (WORKED_Y != 1) 2. Did not work in the past 7 days (WORKED_7 != 1) 3. Was not absent from work (ABSENT != 1) 4. Performed household duties (NOTLAB_WB=1). *Note: Performed household duties should be checked from the responses used to code NOTLAB_WB. If no question can be found to code NOTLAB_WB, use information from the section on household chores*. **Stage 2: Checking for relevant conditions to classify individuals identified above as EMPLHOME** If such an individual belongs to a household that carries out agricultural activities (farming/ sharecropping/ hunting/ fishing/ logging/forestry/raising livestock/bee-keeping), then, such an individual should be considered as employed and coded as 1(Yes), otherwise code as 0(No).
L17	**EMPTYPE_WB_1**	**Type of employment, primary job**	See Appendix I for further details. To generate this variable look for all available information in the following

		1 = Wage & salaried worker 2 = Self-employed with employees 3 =Self-employed w/o employees 4= Employer 5 = Domestic employees (work for compensation for private house) 6 = Family worker 7= Apprentice 9 = volunteer 99=Other/ Not known/ Missing.	sequence: **1.** First use the longest recall period (e.g. 12 month recall period) to code this variable. **2.** Replace the value of missing (coded 99) and volunteer/other (coded 9) based on information provided in sections with lesser recall periods (e.g. 7 day recall period) **3.** Replace the value of the missing (coded 99) and volunteer/other (coded 9) based on information provided in other sections on 'agriculture' and 'enterprise' by looking for explicit individual level information in these sections (i.e. the person responsible or working at the farm/enterprise). Match individual IDs from enterprise/farm sections with those recorded in the individual level file. **4.** Replace the value of the missing (coded 99) and volunteer/other (coded 9) based on information provided at the household or farm or enterprise level. If the household is involved in any agricultural activity, then code the household head as 3 (self-employed without employees). If the household owns a household enterprise, then code the household head as 2 (self-employed with employees or 3 (self-employed without employees) based on number of employees. If no information on the number of employees is available then code as 3 (self-employed without employees). **5.** Replace the value of the missing (coded 99) and volunteer/other (coded 9) based on EMPLHOME/ FETCHWOOD/ FETCHWATER. If EMPLHOME=1 OR FETCHWOOD=1 OR FETCHWATER=1 then code such individuals as 6 (Family worker).

			In questionnaires that do not explicitly ask for primary and secondary jobs (eg. Malawi), ascertain the primary and secondary job by comparing and ranking the jobs in descending order of the time spent in each job. The comparison should be made from all sections (labor, enterprise, agriculture etc.), that ask questions over the past 12 months. Thus the job which takes most work hours is the primary job. However, even while following this methodology the above-mentioned sequence must be maintained, i. e. 12 month questions must be given precedence over 7 day questions and so on. ***Category definitions*** 1) *Wage and salaried* worker is an individual employed by others outside the household including workers who are paid on a daily or hourly basis or based on task. 2) *Self-employed with employees* is an individual who runs a (farm or non-farm) household enterprise or farm that employs at least one non-household member for pay during the reference period. 3) A *self-employed without employees* is an individual who works alone with no paid employees. May have other family members categorized as family or family unpaid workers. May have an apprentice. Sometimes coded as "own account" worker in questionnaires. *4)* An *employer* owns an incorporated business separate from the household, with paid employees. It could also be a large commercial farm. This category is coded only if reported as employer in the questionnaire. 5) *Apprentice* may or may not be paid 6) *Family worker* can be farm or non-farm. Note that family workers may be paid or unpaid. 7) The term *household enterprise* refers to

			a non-farm business or home production activity that is run by household members. It needs not be physically located at the household dwelling unit. It refers to both : a) Businesses engaged in market based transactions and b) Activities designed to produce home consumption.
L18	**EMP_IND_SVY_1**	**Industry classification for primary employment**	Use information from all sections (12 months and 7 days) that ask this question. First use information from the 12 month section. For all missing values and observations coded as 9, use the 7 day section. Keep the original classification of the survey. Most countries follow the ISIC classification.
L19	**INDUSTRY_1**	**1 digit Industry classification for primary employment** *1 = Agriculture and Fishing* *2 = Mining* *3 = Manufacturing* *4 = Electricity and Utilities* *5 = Construction* *6 = Commerce* *7 = Transportation, Storage and Communication* *8 = Financial, Insurance and Real Estate* *9 = Services: Public Administration* *10 = Other Services* *11 = Unspecified*	Use information from all sections (12 months/7 days/any other section) that may provide information. First use information from the 12 month section. For all missing values and observations coded as 9, use the 7 day section. If data allows generation of EMP_IND_SVY_1, recode it to generate INDUSTRY_1 according to SHIP requirements (refer Annexure-2). If data does not follow ISIC classification system, make appropriate assumptions to classify employment industry. Carefully document assumptions. Individuals identified in any section on agricultural activities are coded as 1(Agricultural and fishing). Individuals identified in a household enterprise section and where the industry of the enterprise is available, code them accordingly. In many household enterprise sections, the industry of the enterprise is not available. In such cases do not classify the individual in any industry. Heads of households engaged in agricultural

			activities are coded as 1 (Agriculture and Fishing) Heads of households owning household enterprises are coded according to the industry of the household enterprise, if this information is available; or left as un-coded if the industry of the enterprise is not available. **Individuals identified as EMPLHOME=1 OR FETCHWOOD=1 OR FETCHWATER=1 are coded as 1 (Agriculture and fishing).**
L20	**EMPFRM_1**	**Farm employment in primary job** *0 = No* *1= Yes*	Farming activities include cultivating, sharecropping, forestry, raising livestock, bee-keeping, fishing and hunting. **If INDUSTRY_1=1 then code as 1 (YES).** **If INDUSTRY_1=2 to 10 then code as 0 (No)** If INDUSTRY_1 variable is missing or 11 (unspecified) , code EMPFRM_1 on the basis of all information available in any section of the survey. Individuals identified in household non-agricultural enterprises should be coded as 0(No). These individuals may have missing values for INDUSTRY_1, if the industry information of the enterprise was not available. Similarly heads of households owning non-agricultural enterprises should be coded as 0 (NO). As far as possible, EMPFRM_1 should be coded for all individuals coded as 1 to 9 in EMPTYPE_WB_1. Less than one percent of missing values are allowed for EMPFRM_1 if EMPTYPE_WB_1 has a meaningful value (i.e. NOT 99 (missing)).
L21	**EMPSEC_1**	**Employment sector of primary job** *1 = Public Government, Army* *2 = State Owned* *3 = Private or NGO*	**Code only for EMPTYPE_WB_1=1, 4, 5, 7, 9 (wage employees), otherwise code as missing.** Classifies the main job's sector of activity for individuals with wage jobs. Use information from all sections (12 months and 7 days) that ask this question. First use information from the 12 month section. Then, for all missing values and

			observations coded as 9, use the 7 day section. *Armed forces and international institutions are included in Public sector*. *State owned* includes para-statal firms and all other enterprises in which the government has control. ***Note: if no such questions, leave as missing. This will mean that in the variable EMPCAT_1 wage jobs will not be further disaggregated. Do not code it on the basis of ISCO or ISIC codes.***
L22	**CONTRACT**	**Employment contract** *0 = No* *1 = Yes*	**Code only for EMPTYPE_WB_1=1, 4, 5, 7, 9 (wage employees), otherwise code as missing.** Indicates if a person has a signed (formal) contract, regardless of the duration. This variable is constructed only if there is an explicit question about contracts otherwise code as missing. Use information from all sections (12 months and 7 days) that ask this question. First use information from the 12 month section. Then for all missing values and observations coded as 9, use the 7 day section.
L23	**EMP_CAT_1**	**Employment categories for primary job** *Wage worker* 11=wage public 12=wage private non-agriculture 13=wage private agriculture *Self-employment/ nonwage* 21=family enterprise non-agri. 22=family farmer **31=Not known**	**This is a derived variable from EMPTYPE_WB _1, EMPSEC_1 and EMPFRM_1.** **If EMPTYPE_WB _1=1, 4, 5, 7, 9 then code as 1** If EMP_CAT_1=1 and EMPSEC_1=1, 2 then code 11 If EMP_CAT_1=1 and EMPSEC_1= 3 and EMPFRM_1!=1 then code 12 If EMP_CAT_1=1 and EMPSEC_1= 3 and EMPFRM_1=1 then code 13 **If EMPTYPE_WB_1=2, 3, 6 then code as 2** If EMP_CAT_1=2 and EMPFRM_1!=1 then code 21 If EMP_CAT_1=2 and EMPFRM_1=1 then

			code 22 **If EMPTYPE_WB_1=10 then code as 31**
L24	**WAGE_1**	**Last wage payment in nominal local currency for primary employment**	**For EMPTYPE_WB_1=1, 4, 5, 7, 9 only (wage employees), otherwise code as missing.** Use information from all sections (12 months and 7 days) that ask this question. First use information from the 12 month section. Then, for observations coded as 9 and all missing values, use the 7 day section. Wage from main job (job to which the person dedicated most time in the week preceding the survey). Note: This **includes tips, bonuses, and other cash payments**. **Do not include** the **monetary value** of **in-kind payments** or **perks** associated with the job By definition non-paid employees should have wage=0.
L25	**UNITWAGE_1**	**Time unit for wage in primary employment** *1 = Hourly* *2 = Daily* *3 = Weekly* *4 = Every two weeks* *5 = Twice per month* *6 = Monthly* *7 = Quarterly* *8 = Biannual* *9 = Annually* *10 = Other*	**Code only for EMPTYPE_WB_1=1, 4, 5, 7, 9 (wage employees), otherwise code as missing.** Unit of time for the wage variable of the **main** job. Use information from all sections (12 months and 7 days) that ask this question. First use information from the 12 month section. Then, for all missing values and observations coded as 9, use the 7 day section.
L26	**WHOURS_1**	**Hours worked per week in primary employment**	This is the usual number of hours worked per week in the **main** job. Use information from all sections (12 months and 7 days) that ask this question. First use information from the 12 month section. Then, for all missing values and observations coded as 9, use the 7 day

			section. If only information on how many hours individuals worked per day is available, but no information on number of days worked per week is available, multiply the hours by 5 days. Similarly if hours worked per month is available, divide by 4.2 to get weekly hours.
L27	**SOU_AGRI_1**	**Source of individual income of primary employment: From self-employed agricultural job?** *0 = No* *1 = Yes*	Code as 1(YES) if the source of income is from self-employed agriculture activities, otherwise code as 0 (NO). **If EMP_CAT_1! =. Then code as 0** **Replace code = 1 If EMP_CAT_1=13 or 22**
L28	**SOU_NFARM_1**	**Source of individual income of primary employment: From self-employed non-agricultural job?** *0 = No* *1 = Yes*	Code as 1(YES) if the source of income is from self-employed, non-farming activities, otherwise code as 0 (NO). **If EMP_CAT_1! =. Then code as 0** **Replace code = 1 If EMP_CAT_1=21**
L29	**SOU_WAGE_1**	**Source of individual income of primary employment: From non-agricultural wage job?** *0 = No* *1 = Yes*	Code as 1(YES) if the source of income is from non-agricultural wage job, otherwise code as 0 (NO). **If EMP_CAT_1! =. Then code as 0** **Replace code = 1 If EMP_CAT_1=11 or 12**
L30	**UNEMPLOYED_ILO**	**Unemployment ILO** *0 = No* *1 = Yes*	Individuals are considered as UNEMPLOYED if they meet **ALL** the following conditions: **1.** Did not work in the last 7 days (WORKED_7!=1) **and** **2.** Were not absent from their job (ABSENT !=1) **and** **3.** Looked for a job (LOOKJOB=1) Code as 0 (NO) if WORKED_7=1 **OR** ABSENT=1 To exclude 'employed individuals who are engaged in household duties from unemployed please replace: **UNEMPLOYED_ILO=0 if EMPLHOME==1**
L31	**EMPLOYED_WB**	**Employment WB** *0 = No* *1 = Yes*	This is a derived employment status that has the most inclusive definition.

			If LABFORCE_WB=1 then code 0; Replace code with 1 (YES) if WORKED_Y=1 **or** WORKED_7=1 **or** EMPLHOME=1 **or** ABSENT = 1 **or** EMPTYPE_WB_1! = 99. Replace code with 0 (NO) if EMPTYPE_WB_1 = 99. (If an individual worked in the last year **or** last 7 days **or** had a job to go back to **or** is an individual who fetched water/firewood/ **or** worked on family farm or raised livestock; then such an individual is considered employed.) **Note: Do not miss-assign 0 to missing values. 0 should be coded if ANY of the variables use to derive this variable have non missing values. If ALL the values of the deriving variables are missing, then this variable should also be missing.**
L32	**UNEMPLOYED_WB**	**Unemployment WB** *0 = No* *1 = Yes*	This is a derived unemployment status that has the most exclusive definition. If LABFORCE_WB =1 then code 0; Replace code with 1 (YES) if WORKED_Y! =1 and UNEMPLOYED_ILO = 1 and EMPTYPE_WB_1 = 99. Replace code with 0 (NO) if EMPTYPE_WB_1! = 99. (If an individual did not work in past one year **and** did not work in the past seven days **and** did not work on fetching water, firewood, etc. **and** looked for a job; then such an individual is considered unemployed). **Note: Do not miss-assign 0 to missing values. 0 should be coded if ANY of the variables use to derive this variable have non missing values. If ALL the values of the deriving variables are missing, then this variable should also be missing.**
L33	**LABFORCE_WB**	**Labor-force Participation WB** *0 = No* *1 = Yes*	This is a derived employment status that has the most inclusive definition. First Code 0 (NO) if WORKED_Y ==0 **or** EMPLOYED_7==0 **or** EMPLHOME==0 **or** ABSENT ==0 **or** UNEMPLOYED_ILO==0

			Replace with 1(YES) if WORKED_Y=1 **or** WORKED_7=1 **or** EMPLHOME=1 **or** ABSENT = 1 **or** UNEMPLOYED_ILO=1 **or** EMPTYPE_WB_1! =99. (If an individual worked for last year **or** last 7 days **or** had a job to go back to **or** an individual fetched water/firewood/ **or** worked on family farm or raised livestock, **or** is looking for a job; then such an individual is considered to be in the labor force). **Note: Do not miss-assign 0 to missing values. 0 should be coded if ANY of the variables used to derive this variable have non missing values. If ALL the values of the deriving variables are missing, then this variable should also be missing.**
L34	**HOUR_NECON**	**Hours spent per week on non-economic activities.**	One month = 4-weeks by convention. These include activities such as preparing food, and care for children. Fetching wood and water are considered economic activities and should **not** be included here.

TABLE 2.4 SECONDARY EMPLOYMENT (CCD_YEAR_L.DTA)

No	NAME	LABEL AND CODES	COMMENTS, QUALITY CONTROL CHECKS AND CORRECTIONS
L35	**EMPTYPE_WB_2**	**Type of employment Secondary job** *1 = Wage & salaried worker* *2 = Self-employed with employees* *3 =Self-employed w/o employees* *4= Employer* *5 = Domestic employees (work for compensation for private house)* *6 = Family worker* *7= Apprentice* *9 = volunteer* *99 = Other /Not known/ Missing*	To generate this variable only look for information in the following sequence:- **1.** First use the longest recall period (e.g. 12 month recall period) to code this variable **2.** Replace the value of missing (coded 99) and volunteer/other (coded 9) based on information provided in sections with lesser recall periods (e.g. 7 day recall period) Do not look for additional information beyond the secondary job section of the questionnaire. Do not code household heads or EMPLHOME or FETCHWOOD or FETCHWOOD as done in EMPTYPE_WB_1.

L36	**EMP_IND_SVY_2**	**Industry classification secondary employment**	Use information from all sections (12 months and 7 days) that ask this question. First use information from the 12 month section. Then for all missing values and observations coded as other, use the 7 day section. Keep the original classification of the survey. Most countries follow the ISIC classification.
L37	**INDUSTRY_2**	**1 digit Industry classification secondary employment** *Refer Annexure 2 for details* 1 = Agriculture and Fishing 2 = Mining 3 = Manufacturing 4 = Electricity and Utilities 5 = Construction 6 = Commerce 7 = Transportation, Storage and Communication 8 = Financial, Insurance and Real Estate 9 = Services, Public Administration 10 = Other Services 11 = Unspecified	Classifies the **second job** of an individual **with a second job**. Should be missing otherwise. If data allows generation of EMP_IND_SVY_2, recode INDUSTRY_2 according to SHIP requirements (refer Annexure-2) If data does not follow ISIC classification system, make appropriate assumptions to classify employment industry. Carefully document assumptions in .do file. **Do not code further** based on implied information for household heads or EMPLHOME or FETCHWOOD or FETCHWOOD as was done in INDUSTRY_1.
L38	**EMPFRM_2**	**Farm employment in primary job** *0 = No* *1= Yes*	Farming activities include cultivating, sharecropping, forestry, raising livestock, bee-keeping, fishing and hunting. **If INDUSTRY_2=1 then code as 1 (YES).** **If INDUSTRY_2=2 to 10 then code as 0 (No)** **Do not code further** based on implied information for household heads or EMPLHOME or FETCHWOOD or FETCHWOOD as was done in EMPFRM_1.

L39	**EMPSEC_2**	**Employment sector of primary job** 1 = Public Government, Army 2 = State Owned 3 = Private, NGO	**Code only for EMPTYPE_WB_2=1, 4, 5, 7, 9 (wage employees), otherwise code as missing.** Classifies the sector of activity of the secondary job for individuals with wage jobs. Use information from all sections (12 months or 7 days) that ask this question. First use information from the 12 month section. For all missing values and for observations coded as 9, use the 7 day section. *Armed forces and international institutions are included in Public sector*. *State owned* includes parastatal firms and all other enterprises in which the government has control. ***Note: if no such question leave as missing. This will mean that the in the variable EMPCAT_2 wage jobs will not be further disaggregated. Do not code it on the basis of ISCO or ISIC codes.***
L40	**EMP_CAT_2**	**Employment categories for primary job** *1= Wage worker* *11=wage public* *12=wage private non-agriculture* *13=wage private agriculture* *2 = Self-employment/ nonwage* *21=family enterprise non-agri.* *22=family farmer*	**This is a derived variable from the EMPTYPE_WB_2, EMPSEC_2 and EMPFRM_2.** **If EMPTYPE_WB_2=1, 4, 5, 7, 9 then code as 1** If EMP_CAT_2=1 and EMPSEC_2=1, 2 then code 11 If EMP_CAT_2=1 and EMPSEC_2= 3 and EMPFRM_2=0 then code 12 If EMP_CAT_2=1 and EMPSEC_2= 3 and EMPFRM_1=1 then code 13 **If EMPTYPE_WB_2=2, 3, 6 then code as 2** If EMP_CAT_2=2 and EMPFRM_2=0 then code 21 If EMP_CAT_2=2 and EMPFRM_2=1 then code 22
L41	**WAGE_2**	**Last wage payment in nominal local currency for secondary employment**	**For EMPTYPE_WB_1=1, 4, 5, 7, 9 only (wage employees).** Use information from all sections (12 months and 7 days) that ask this question. First use information from the 12 month section. Then, for all missing values and

			observations coded as 9, use the 7 day section. Wage from secondary job. This includes tips, bonuses, and other wage payments. By definition non-paid employees should have wage=0.
L42	**UNITWAGE_2**	**Time unit for wage in secondary employment** 1 = Hourly 2 = Daily 3 = Weekly 4 = Every two weeks 5 = Twice per month 6 = Monthly 7 = Quarterly 8 = Biannual 9 = Annually 10 = Other	**Code only for EMPTYPE_WB_1=1, 4, 5, 7, 9 (wage employees), otherwise code as missing.** Unit of time for the wage variable of the secondary job. Use information from all sections (12 months and 7 days) that ask this question. First use information from the 12 month section. Then for all missing values and observations coded as 9, use the 7 day section.
L43	**WHOURS_2**	**Hours worked in last week in secondary employment**	This is the usual number of hours worked per week in the **secondary** job. Use information from all sections (12 months or 7 days) that ask this question. First use information from the 12 month section. For all missing values, use the 7 day section. If only information on how many hours individuals worked per day is available but no information on number of days worked a week is available, multiply the hours by 5 days. Similarly for hours worked per month divide by 4.2 to get weekly hours.
L44	**SOU_AGRI_2**	**Source of individual income of secondary employment: From self- employed Agricultural job?** *0 = No* *1= Yes*	Code as 1(YES) if the source of income is from self-employed agriculture activities, otherwise code as 0 (NO). If EMP_CAT_2! =. OR EMP_CAT_2! =. Then code as 0 Replace code = 1 If EMP_CAT_2=13 or 22
L45	**SOU_NFARM_2**	**Source of individual inc of 2nd employment: from self-employed non-agri job?** *0 = No* *1= Yes*	Code as 1(YES) if the source of income is from self-employed agriculture activities, otherwise code as 0 (NO). If EMP_CAT_2! =. Then code as 0 Replace code = 1 If EMP_CAT_2=21

L46	**SOU_WAGE_2**	**Source of individual income of secondary employment: From a non-agricultural wage job?** *0 = No* *1= Yes*	Code as 1(YES) if the source of income is from self-employed agriculture activities, otherwise code as 0 (NO). If EMP_CAT_2! =. Then code as 0 Replace code = 1 If EMP_CAT_2=11 or 12

CHAPTER 3: MODULE III – HARMONIZED HOUSEHOLD-LEVEL VARIABLES

(All variables in this chapter to be included in the CCD_YEAR_H.dta)

This module records household-level information and includes household head characteristics and some selected information from individual-level module. Other information include housing characteristics and utilities, access to various amenities measured in terms of distances/time and ownership of durable goods among others. All variables are numeric unless otherwise stated.

TABLE 1.0 SAMPLE, GEOGRAPHY AND BASIC HOUSEHOLD IDENTIFIER (CCD_YEAR_H.DTA)

No	SHIP VARIABLE NAME	LABEL AND CODES	COMMENTS, QUALITY CONTROL CHECKS AND CORRECTIONS
1	**COUNTRY**	**Country code**	To be merged from CCD_YEAR_BASICINFO.dta
2	**REGION**	**Region code and label the names**	To be merged from CCD_YEAR_BASICINFO.dta
2b	**PROVINCE**	**Province code and label the names**	To be merged from CCD_YEAR_BASICINFO.dta
2c	**DISTRICT**	**District code and label the names (if available)**	To be merged from CCD_YEAR_BASICINFO.dta
3	**HID**	**Household unique identification**	To be merged from CCD_YEAR_BASICINFO.dta
4a	**SURVEMO**	**Month of survey**	To be merged from CCD_YEAR_BASICINFO.dta
4b	**SURVEYR**	**Year of survey**	To be merged from CCD_YEAR_BASICINFO.dta
5	**RURURB**	**Area of residence** *1 = Rural* *2 = Urban*	To be merged from CCD_YEAR_BASICINFO.dta
6a	**HHSIZE**	**Household members**	To be merged from CCD_YEAR_BASICINFO.dta
6b	**HHSIZE_S**	**Household members including servants**	To be merged from CCD_YEAR_BASICINFO.dta
7	**WTA_S**	**Individual weighting coefficient**	To be merged from CCD_YEAR_BASICINFO.dta
8	**WTA_S_HHSIZE**	**Household weighting coefficient** WTA_S_HHSIZE=WTA_S * HHSIZE	To be merged from CCD_YEAR_BASICINFO.dta

TABLE 3.1 HOUSEHOLD HEAD CHARACTERISTICS (CCD_YEAR_H.DTA)

No	NAME	LABEL and CODES	COMMENTS, QUALITY CONTROL CHECKS AND CORRECTIONS
H1	**HHSEX**	**Sex of household Head** *1 = Male* *0 = Female*	To be merged from CCD_YEAR_I.dta "SEX".
H2	**HHAGEY**	**Age of household Head**	To be merged from CCD_YEAR_I.dta "AGEY".
H3	**HHMARST**	**Marital status of household Head** *1 = Never married* *2 = Married monogamous* *3 = Married polygamous* *4 =Common law, union coutumiere, union libre, living together* *5 = Divorced/Separated* *6 = Widowed*	To be merged CCD_YEAR_I.dta "MARSTAT".
H4	**HHEDLEV**	**Education level of household Head** *0 = No education* *1 = Pre-school* *2 = Primary, not completed* *3 = Completed primary, but less than completed lower secondary* *4 = Completed lower secondary (or post-primary vocational education) but less than completed upper secondary* *5 = Completed upper secondary (or extended vocational/technical education)* *6 = Post-secondary technical* *7 = University and higher* *8 = Formal adult education or literacy program* *9 = Other*	To be merged from CCD_YEAR_I.dta "EDLEVEL_AR.
H5	**HHEMSTAT**	**Employment status of household Head (usual)** *0 = Unemployed* *1 = Employed*	To be merged from CCD_YEAR_L.dta. Use variable EMPLOYED_WB and UNEMPLOYED_WB from Labor module.
H6	**HHEMTYPE**	**Employment type of household Head** *1 = Wage & salaried worker*	To be merged from CCD_YEAR_L.dta. For primary job of household head only.

		2 = Self-employed with employees *3 =Self-employed w/o employees* *4= Employer* *5 = Domestic employees (work for compensation for private house)* *6 = Family worker* *7= Apprentice* *9 = volunteer/Other*	Use variable EMPTYPE_WB_1 from CCD_YEAR_L.dta
H7	**HHINDUSTRY**	**Sector of activity of household Head** *1 = Agriculture and fishing* *2 = Mining* *3 = Manufacturing* *4 = Electricity and utilities* *5 = Construction* *6 = Commerce* *7 = Transport, storage and communication* *8 = Financial, insurance and real estate* *9 = Services; Public administration* *10 = Other services* *11 = Unspecified*	To be merged from CCD_YEAR_L.dta. For the primary job of household head only. Use variable INDUSTRY_1 from Labor module

TABLE 3.2 HOUSING AND UTILITIES (CCD_YEAR_H.DTA)

No	NAME	LABEL and CODES	COMMENTS, QUALITY CONTROL CHECKS and CORRECTIONS
H8	OWNHOUSE	**Ownership of dwelling unit** *1 = Yes* *0 = No*	YES refers to: (a) Actual ownership or (b) If household is in the process of purchasing the dwelling unit.
H9	ROOMS	**Number of habitable rooms**	This refers to number of rooms in the dwelling unit and may consist of one or more structures. Must be >= 1. Includes all rooms used for living, sleeping and eating. Excludes store, bathrooms and kitchens.
H 10	ROOF	**Main material used for roof** *1 = Concrete/cement/ brick/stone* *2 = Wood* *3 = Bamboo/thatch*	

		4 = Tiles/shingles *5 = Tin/metal sheets* *6 = Makeshift* *9 = Other*	
H 11	WALLS	**Main material used for external walls** *1 = Concrete/cement/ brick/stone* *2 = Wood* *3 = Bamboo/thatch* *4 = Iron/metal sheets* *5 = Clay/mud* *6 = Makeshift* *9 = Other*	
H 12	FLOOR	**Main material used for floor** *1 = Cement/tiles/marble* *2 = Wood/Bamboo* *3 = Vinyl* *4 = Earth/clay/mud* *9 = Other*	
H 13	WATER	**Main source of water** *1 = Pipe (own tap)* *2 = Public standpipe* *3 = Borehole* *4 = Wells (protected)* *5 = Wells (unprotected)* *6 = Surface water* *7 = Rain water* *8 = Vendor/truck* *9 = Other*	Drinking water if specified.
H 14	ADISWAT	**Actual distance to main water point (kms)**	This refers to actual distance to water point (***one way***) used by household in kms. 1km = 1000 m; 1 km = 5/8 mile. If within dwelling, zero
H 15	ATIMWAT	**Actual time taken to main water point (mins)**	This refers to actual time taken to water point used by household. If roundtrip provided, divide by 2.
H 16	ELECTCON	**Connection of electricity in dwelling** *1 = Central/local* *2 = Solar/wind* *3 = Generator* *9 = Other* *0 = None*	If electricity source connection cannot be determined code 9 (*Yes, but source unstated*).
H 17	FUELCOOK	**Main cooking fuel** *1 = Firewood* *2 = Kerosene*	*Other* includes fuel derived from coffee waste, saw dust, crop residue, cow dung among others.

		3 = Charcoal *4 = Electricity* *5 = Gas* *9 = Other*	
H 18	FUELLIGH	**Main lighting fuel** *1 = Electricity* *2 = Kerosene* *3 = Candles* *4 = Gas* *9 = Other*	
H 19	TOILET	**Main toilet facility** *1 = Flush toilet* *2 = Improved pit latrine* *3 = Pit latrine* *4 = No facility* *9 = Other*	
H 20	GARBDISP	**Garbage and trash disposal** *1 = Collected* *2 = Buried/burned* *3 = Discarded in empty lots, street, rivers* *9 = Other*	Refers to only garbage or trash generated by household.
H 21	DISPSCH	**Distance to nearest elementary/primary school (kms)**	This refers to one-way distance to nearest primary school in kms.
H 22	DISPSCH1	**Distance to nearest elementary/primary school** *1 = Outside dwelling and up to 1 km* *2 = >1-2 km* *3 = >2-5 km* *4 = >5-10 km* *5 = Over 10 km* *9 = Undefined*	One-way
H 23	TIMPSCH	**Time taken to nearest elementary/primary school (minutes)**	One way
H 24	TIMPSCH1	**Time taken to nearest elementary/primary school** *1 = Within ½ hour* *2 = >½ to 1 hour away* *3 = >1 to 6 hours* *4 = >6 to 12 hrs* *5 = Over 12 hrs* *9 = Undefined*	One way
H 25	DISHEAL	**Distance to nearest health facility (kms)**	One way
H 26	DISHEAL1	**Distance to nearest health**	One way

		facility *1 = Outside dwelling and up to 1 km* *2 = >1-2 km* *3 = >2-5 km* *4 = >5-10 km* *5 = Over 10 km* *9 = Undefined*	
H 27	TIMHEAL	**Time taken to nearest health facility (minutes)**	One way
H 28	TIMHEAL1	**Time taken to nearest health facility** *1 = Within ½ hour* *2 = >½ to 1 hour away* *3 = >1 to 6 hours* *4 = >6 to 12 hrs* *5 = Over 12 hrs* *9 = Undefined*	One way

TABLE 3.3 OWNERSHIP OF DURABLE AND NON-DURABLE GOODS (CCD_YEAR_H.DTA)

No	NAME	LABEL AND CODES	COMMENTS, QUALITY CONTROL CHECKS AND CORRECTIONS
H 29	**AGLAND**	**Ownership of agricultural land** *1 = Yes* *2 = Cannot be determined* *3 = No*	Country specific on definition of land ownership. This only refers to ***agricultural land***. Some countries inquire about land ownership irrespective whether it is agricultural or non-agricultural purposes. If the survey is not explicit on type of land code 2 and skip to LANDSIZE.
H 30	**AGLNDOW**	**Agricultural land owned by household (ha)**	Refers only to **agricultural** land size in hectares. Check, if AGLAND==3 , code as 0 By convention 1 ha = 2.471 acres. **Note: This would exclude rental farmers but include farm land owners who are not farmers.**
H 31	**OTHLNDOW**	**Non Agricultural land owned by household (ha)**	Refers only to **non-agricultural** land size in hectares. By convention 1 ha = 2.471 acres.
H 32	**LANDSIZE**	**Land size owned by household (ha)**	Area of all land owned by household regardless of use, that is, both agricultural and non-agricultural. By convention 1 ha = 2.471 acres.

H 33	**AREA_IRRIG**	**Area of irrigated agricultural land (ha)**	
H 34	**AREA_NONIRRIG**	**Area of non-irrigated agricultural land (ha)**	
H 35	**AREA_COLLAT**	**Total plot areas can be used as collateral for loan (ha)**	
H 36	**LANDCON**	**Conflict about the land** *1 = Yes* *0 = No*	
H 37	**RADIO**	**Ownership of radio** *1 = Yes* *0 = No*	Functioning radio includes a radio, radio cassette, and 3-in-1-radio cassette.
H 38	**TV**	**Ownership of television** *1 = Yes* *0 = No*	Presence of a functioning television in house.
H 39	**PHONE**	**Ownership of landline (fixed) phone** *1 = Yes* *0 = No*	Presence of a functioning fixed land line telephone in house.
H 40	**CPHONE**	**Ownership of at least one cellular phone** *1 = Yes* *0 = No*	Presence of a functioning cellular in house.
H 41	**INTERNET**	**Access to internet** *1 = Yes* *0 = No*	Include either inside or outside the house.
H 42	**FRIDGE**	**Ownership of frig** *1 = Yes* *0 = No*	Presence of a functioning refrigerator in house.
H 43	**SEWMACH**	**Ownership of sewing machine** *1 = Yes* *0 = No*	Presence of a functioning sewing machine in house.
H 44	**COMPUTER**	**Ownership of computer** *1 = Yes* *0 = No*	Presence of a functioning computer in house.
H 45	**STOVE**	**Ownership of stove** *1 = Yes* *0 = No*	Presence of a functioning stove or cooker in house.
H 46	**OXCART**	**Ownership of animal cart** *1 = Yes* *0 = No*	Presence of a functioning animal cart, which is used as a means of transport or a farm tool.
H 47	**BCYCLE**	**Ownership of bicycle** *1 = Yes* *0 = No*	Presence of a functioning bicycle.
H 48	**BOAT**	**Ownership of boat** *1 = Yes* *0 = No*	Presence of a functioning boat or canoe.

H 49	**MCYCLE**	**Ownership of motorcycle** *1 = Yes* *0 = No*	Presence of a functioning motorcycle is important.
H 50	**CAR**	**Ownership of private car** *1 = Yes* *0 = No*	Presence of a functioning car is important. This refers to car for household use and NOT a commercial vehicle.
H 51	**LLIVESK**	**Number of large-sized livestock owned**	This question asks the actual number of large-sized livestock. These include cattle, camels, donkeys and horse.
H 52	**MLIVESK**	**Number of medium-sized livestock owned**	This question asks the actual number of medium-sized livestock. These include sheep, goats and pigs.
H 53	**POULTRY**	**Number of poultry owned**	This question asks the actual number of poultry/birds. These include all forms of birds such as chicken, geese, and doves.

TABLE 3.4 HOUSEHOLD EXPENDITURE (CCD_YEAR_H.DTA)

H 54	**PCEXP**	**Per capita annual food and non-food consumption expenditure in national currency, not regionally deflated**	To be merged from CCD_YEAR_E.dta.
H 55	**PCEXPDR**	**Per capita annual food and non-food consumption expenditure in national currency, adjusted by regional food price indices, and *cleaned.***	To be merged from CCD_YEAR_E.dta Per capital annual food and non-food consumption is first adjusted by regional food price indices. Then outliers (separately for food and non-food expenditure) greater than three standard deviations are replaced with the medians of their correspondent per capita food and non-food consumption by region and rural/urban.
H 56	**PCEXPDR_PPP**	**Per capita annual food and non-food consumption expenditure in real terms, converted to 2005 $PPP using 2005 $PPP exchange rate**	To be merged from CCD_YEAR_E.dta Calculated using PCEXPDR, CPI of the survey year, and 2005 $PPP conversion factor. When the regional food price deflators are not available, remove the outliers in PCEXP following the same procedures as removing the outliers in PCEXPDR, and calculate PCEXP_PPP using PCEXP, CPI of the survey year, and 2005 $PPP conversion factor.
H 57	**QUINTILE_N**	**Income quintiles at the national level**	Calculated based on PCEXPDR_PPP, and weighted by population weights.
H 58	**QUINTILE_RU**	**Income quintiles by rural/urban**	Calculated based on PCEXPDR_PPP, and weighted by population weights.

TABLE 3.5 HOUSEHOLD REMITTANCES (CCD_YEAR_H.DTA)

H 59a	**SEX_RMT_1**	**Sex of the 1st, 2nd, and 3rd remittance sending member** *1 = Male* *0 = Female*	The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).
H 59b	**SEX_RMT_2**		
H 59c	**SEX_RMT_3**		
H 60a	**RELAT_RMT_1**	**Relationship to the household head of the 1st, 2nd, and 3rd remittance sending member** *2 = Spouse* *3 = Son/daughter* *4 = Parents/grandparents* *5 = Grandchild* *6 = Son/daughter-in-law* *7 = Siblings and other relative*	The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).
H 60b	**RELAT_RMT_2**		
H 60c	**RELAT_RMT_3**		
H 61a	**DES_MIG_1**	**Destination of migration of the 1st, 2nd, and 3rd remittance sending member** *1 = Capital* *2 = Within the country (but not capital)* *3 = Abroad*	The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).
H 61b	**DES_MIG_2**		
H 61c	**DES_MIG_3**		
H 62	**ORIGIN_RMT**	**Origin of the remittances** *1 = Domestic* *2 = Abroad* *3 = Both*	Location category from where the household has received remittances.
H 63a	**AMT_RMT_1**	**Amount of annual remittance by the 1st, 2nd, and 3rd remittance sending member**	The order of the sending members is in decreasing order of amount of remittance (remittance includes cash, gifts and food).
H 63b	**AMT_RMT_2**		
H 63c	**AMT_RMT_3**		
H 64	**INC_RMT**	**Total amount of annual remittances received from remittance sending members (annual)**	The total includes the remittances from the most important three migrant members and others.

TABLE 3.6 HOUSEHOLD INCOME VARIABLES (CCD_YEAR_H.DTA)

H 65	**INC_WAGE**	**Wage and salaries (annual)**	Aggregated from individual data.
H 66	**INC_SOCA**	**Bonus and social allowances derived from wage jobs (annual)**	Aggregated from individual data.
H 67	**INC_FARM_G**	**Gross income from household farm businesses (annual)**	Gross receipt/revenue. Income from leasing agricultural land is counted in INC_RENT.
H 68	**INC_NFARM_G**	**Gross income from household nonfarm businesses (annual)**	Gross receipt/revenue. Gross receipt/revenue minus operating cost (as reported).
H 69	**INC_RENT**	**Rental income (annual)**	Income from selling or leasing land, houses, and other properties. Most survey questionnaires have asked about both rent from leasing land and houses in the housing module, and rent from leasing land in the agricultural module. Count both, assuming there's no overlap.
H 70	**INC_FIN**	**Financial income (annual)**	Income from savings, dividends, loans, tax refunds, and maturity payment on insurance
H 71	**INC_PEN**	**Pensions/ social assistance (annual)**	Income from pensions and other social assistance. A similar variable is in the labor module. However, that is at an individual level whereas this is at a household level.
H 72	**INC_STUDY**	**Study support (annual)**	Income from scholarships, stipends, or other study support
H 73	**INC_OTH**	**Other income (annual)**	
H 74	**INC_TOT_G**	**Total gross household income, (annual)**	In national currency. Total of labor income and non-labor incomes including financial income, remittances, pensions and other social assistance, study support and other. **Sum of all variables starting with INC_. (Include INC_RMT from section 3.6)**

Note:

a) Any variable, for which information is not collected by the country, should be created and left as system missing in the final harmonized aggregated expenditure file.

b) The order of variables should follow as shown above for ease in reference use.

CHAPTER 4: MODULE IV – HARMONIZED EXPENDITURE VARIABLES

The following guidelines should be kept in mind:

a) Before starting the expenditure aggregation, household size must be first calculated. This is because using expenditure per capita one can determine outliers.
b) If a particular sub-category does not exist in the survey data, list all values as missing.
c) All expenditures should be annualized.
d) Although the summary tables are produced with great care to make the sub-aggregates as well defined and exhaustive as possible, the users will still need to make circumstantial (important) decisions. Thus all computer programs should be attached so the aggregation can be appropriately documented. [2]

The following 14 expenditure files (File 1 to File 14) must be generated, which serve as the record of processing expenditure aggregation and ensure the replicability of the SHIP variables from the original survey data.

Contents	**Sequence**	**Table**	**SHIP Output Data File**
Household composition	File 0	Table0	00_BASICINFO.dta
Food and non-alcoholic beverages	File 1	Table 4.1	01_EXPFOOD.dta
Alcoholic beverages, tobacco and narcotics	File 2	Table 4.2	02_EXPALCH.dta
Clothing and footwear	File 3	Table 4.3	03_EXPCLTH.dta
Housing, water, electricity, gas and other fuels	File 4	Table 4.4	04_EXPHOUS.dta
Furnishings, household equipment, etc.	File 5	Table 4.5	05_EXPFURN.dta
Health	File 6	Table 4.6	06_EXPHLTH.dta
Transport	File 7	Table 4.7	07_EXPTRSP.dta
Communication	File 8	Table 4.8	08_EXPRCRE.dta
Recreation and culture	File 9	Table 4.9	09_EXPCMNQ.dta
Education	File 10	Table 4.10	10_EXPEDUC.dta
Restaurants and hotels	File 11	Table 4.11	11_EXPHOTL.dta
Miscellaneous goods and services	File 12	Table 4.12	12_EXPMISC.dta
Regional food price deflators	File 13	Table 4.13	13_FDPINDEX.dta
Total household expenditures	File 14	Table 4.14	CCD_YEAR_E.dta

[2] Users can also reference other literature on methodology of compiling consumption aggregates. Deaton, Angus and Salman Zaidi (2002), Guidelines for Constructing Consumption Aggregates for Welfare Analysis, Living Standards Measurements Study Working Paper Number 135, The World Bank: Washington DC.

TABLE 1.0: HOUSEHOLD COMPOSITION AND WEIGHTS SUMMARY TABLE

	Variable name	***Variable definition***	***From***
1	COUNTRY	Country code	00_BASICINFO.dta
2	REGION	Region code and label the names	00_BASIC INFO.dta
3	HID	Household unique identification	00_BASIC INFO.dta
4a	SURVEMO	Month of survey	00_BASIC INFO.dta
4b	SURVEYR	Year of survey	00_BASIC INFO.dta
5	RURURB	Area of residence *1 = Rural* *2 = Urban*	00_BASIC INFO.dta
6a	HHSIZE	Household members	00_BASIC INFO.dta
6b	HHSIZE_S	Household members including servants	00_BASIC INFO.dta
7	WTA_S	Individual weighting coefficient	00_BASIC INFO.dta
8	WTA_S_HHSIZE	Household weighting coefficient WTA_S_HHSIZE=WTA_S * HHSIZE	00_BASIC INFO.dta

TABLE 4.1: FOOD EXPENDITURES

All variables are annualized and in current local nominal prices

	Variable name	***Variable definition***	***SHIP output data file***
3	HID	Household unique ID	01_EXPFOOD.dta
E1	FDRECALL	Food purchase recall period, in weeks. By convention 1 week = 7 days.	01_EXPFOOD.dta
E2	FD_B	Total expenditure on purchased food. Beverages, tobacco, catering and restaurant foods are excluded	01_EXPFOOD.dta
E3	FDNONALC_B	Total expenditure on purchased non-alcoholic beverages	01_EXPFOOD.dta
E4	FD_P	Total value of food other than purchased, such as self-produced, in-kind receipts, etc., beverages and tobacco excluded	01_EXPFOOD.dta
E5	FDNONALC_P	Total value of non-alcoholic beverages other than purchased, such as self-produced, in-kind receipts, etc.	01_EXPFOOD.dta
E6	TOTFOOD	Total food expenditure, sum of variables with prefix FD (*Exclude FDRECALL*)	01_EXPFOOD.dta

TABLE 4.2: ALCOHOLIC BEVERAGES, TOBACCO AND NARCOTICS

All variables are annualized and in current local nominal prices

	Variable name	***Variable definition***	***SHIP output data file***
3	HID	Household unique ID	02_EXPALCH.dta
E7	FDALC_B	Total expenditure on purchased alcoholic beverages	02_EXPALCH.dta
E8	FDALC_P	Total value of alcoholic beverages other than purchased, such as self-produced, gifts received, etc.	02_EXPALCH.dta
E9	TOBACCO	Sum of total expenditure on purchased tobacco and tobacco products, and total value of tobacco received in-kind or self-produced	02_EXPALCH.dta
E10	NARCOTICS	Sum of total expenditure on purchased narcotics and narcotics products, and total value of narcotics received in-kind or self-produced	02_EXPALCH.dta
E11	TOTALCH	Total expenditures on alcoholic beverages, tobacco and narcotics, sum of all above variables	02_EXPALCH.dta

TABLE 4.3: CLOTHING AND FOOTWEAR

All variables are annualized and in current local nominal prices

	Variable name	***Variable definition***	***SHIP output data file***
3	HID	Household unique ID	03_EXPCLTH.dta
E12	HSCLOTH	Total expenditure on clothing, including self-produced and in-kind receipts	03_EXPCLTH.dta
E13	HSFOOTW	Total expenditure on footwear, including self-produced and in-kind receipts	03_EXPCLTH.dta
E14	TOTCLTH	Total expenditure on clothing and footwear, including self-produced and in-kind receipts, sum of the above two variables	03_EXPCLTH.dta

TABLE 4.4: HOUSING, WATER, ELECTRICITY, GAS AND OTHER FUELS

For ship files, rent is not included in the final total household expenditure because in SSA normally over 80 percent of households live in their own houses and do not pay mortgage. There is no rental market in rural areas where majority population lives. However, self-reported actual rent variable should be in this file if the rent question is asked. The user can also impute rent for households who do not pay rent or mortgage, which however, is not required by SHIP file.

All variables are annualized and in current local nominal prices

	Variable name	***Variable definition***	***Data files***
3	HID	Household unique ID	04_EXPHOUS.dta
E15	RENT_ACT	Actual rentals for housing	04_EXPHOUS.dta

E16	RENT_IMP	Imputed housing for rental	04_EXPHOUS.dta
E17	HSREPAIR	Maintenance and repair of dwelling	04_EXPHOUS.dta
E18	HSWATER	Water supply and miscellaneous services relating to the dwelling such as garbage collection fees	04_EXPHOUS.dta
E19	HSFUEL	Firewood, and charcoal	04_EXPHOUS.dta
E20	GAS	Gas for cooking	04_EXPHOUS.dta
E21	ELEC	Electricity	04_EXPHOUS.dta
E22	HSKEROSENE	Kerosene (Paraffin)	04_EXPHOUS.dta
E23	HSDIESEL	Diesel for all purposes except transportation. Note: If in the questionnaire does not specify the purpose of diesel use, include it in TRFUEL in Transport Expenditures in Table 4.7.	04_EXPHOUS.dta
E24	HSUTILITY	Electricity and gas (row total of GAS and ELEC)	04_EXPHOUS.dta
E25	TOTHOUS	Total expenditures on housing, sum of variables with the prefix HS	04_EXPHOUS.dta

TABLE 4.5: FURNISHINGS, HOUSEHOLD EQUIPMENT AND ROUTINE HOUSEHOLD MAINTENANCE

All variables are annualized and in current local nominal prices

	Variable name	***Variable definition***	***SHIP output data file***
3	HID	Household unique ID	05_EXPFURN.dta
E26	HSFURNI	Furniture and furnishings, carpets and other floor coverings.	05_EXPFURN.dta
E27	HSTEXTIL	Household textiles such as beddings, and drapes.	05_EXPFURN.dta
E28	HSAPPLIAN	Household appliances	05_EXPFURN.dta
E29	HSUTENSIL	Glassware, tableware, kitchenware, and household utensils.	05_EXPFURN.dta
E30	HSTOOLS	Tools and equipment for house and garden	05_EXPFURN.dta
E31	HSSERV	Goods and services for routine household maintenance, e.g., cleaning products, domestic services.	05_EXPFURN.dta
E32	TOTFURN	Total furnishing etc. expenditure, sum of all variables with the prefix HS	05_EXPFURN.dta

TABLE 4.6: HEALTH CARE EXPENDITURES

All variables are annualized and in current local nominal prices

	Variable name	***Variable definition***	***SHIP output data file***
3	HID	Household unique ID	06_EXPHLTH.dta
E33	HLMEDI	Medical products, appliances and equipment	06_EXPHLTH.dta
E34	HLOUTP	Outpatient services	06_EXPHLTH.dta
E35	HOSP	Hospitalization expenditure	06_EXPHLTH.dta
E36	TOTHLTH	Total health expenditures, sum of all variables with the prefix HL (Do **NOT** include **HOSP**)	06_EXPHLTH.dta

TABLE 4.7: TRANSPORT EXPENDITURES

All variables are annualized and in current local nominal prices

	Variable name	*Variable definition*	*Data files*
3	HID	Household unique ID	07_EXPTRSP.dta
E37	VEHICLES	Purchase of vehicles	07_EXPTRSP.dta
E38	TRFUEL	Fuels for transportation, such as petrol, diesel, etc… Note: If in the questionnaire does not specify the purpose of diesel/petrol use, include it in TRFUEL.	07_EXPTRSP.dta
E39	TRPERSON	Operation of personal transport equipment, such as repairs and hire of chauffeurs, but excluding fuel.	07_EXPTRSP.dta
E40	TRSERVE	Transport services, such as public transportation, purchase of airplane tickets, etc.	07_EXPTRSP.dta
E41	TOTTRSP	Total transport expenditures, sum of all variables with the prefix TR (Do **NOT** include **Vehicles**)	07_EXPTRSP.dta

TABLE 4.8: COMMUNICATION EXPENDITURES

All variables are annualized and in current local nominal prices

	Variable name	*Variable definition*	*SHIP output data file*
3	HID	Household unique ID	08_EXPCMNQ.dta
E42	COMPOST	Postal services	08_EXPCMNQ.dta
E43	COMTEL	Telephone and fax equipment	08_EXPCMNQ.dta
E44	COMSERVE	Telephone, fax, and internet services	08_EXPCMNQ.dta
E45	TOTCMNQ	Total communication expenditures, sum of all variables with the prefix COM	08_EXPCMNQ.dta

TABLE 4.9: RECREATION AND CULTURE EXPENDITURES

All variables are annualized and in current local nominal prices

	Variable name	*Variable definition*	*SHIP output data file*
4	HID	Household unique ID	09_EXPRCRE.dta
E46	RCAUDI	Audio-visual, photographic and information processing equipment	09_EXPRCRE.dta
E47	DURRC	Other major durables for recreation and culture	09_EXPRCRE.dta
E48	RCPETS	Other recreational items and equipment, gardens and pets	09_EXPRCRE.dta
E49	RCSERVE	Recreational and cultural services, such as movies, etc.	09_EXPRCRE.dta
E50	RCNEWS	Newspapers, books and stationary of non-educational purpose	09_EXPRCRE.dta
E51	RCHOLID	Package holidays	09_EXPRCRE.dta
E52	TOTRCRE	Total recreation expenditures, sum of all variables with the prefix RC (Do **NOT** include **DURRC**)	09_EXPRCRE.dta

TABLE 4.10: EDUCATION EXPENDITURES

All variables are annualized and in current local nominal prices

	Variable name	*Variable definition*	*SHIP Output Data files*
3	HID	Household unique ID	10_EXPEDUC.dta
E53	EDPRIMAR	Pre-primary and primary	10_EXPEDUC.dta
E54	EDSECOND	Secondary education	10_EXPEDUC.dta
E55	EDPOSTSEC	Post-secondary but non-tertiary education	10_EXPEDUC.dta
E56	EDTERTIAR	Tertiary education	10_EXPEDUC.dta
E57	EDUNDEFIN	Education cannot be defined	10_EXPEDUC.dta
E58	TOTEDUC	Total expenditure on education, sum of all variables with ED prefix	10_EXPEDUC.dta

TABLE 4.11: RESTAURANTS AND HOTELS

All variables are annualized and in current local nominal prices

	Variable name	*Variable definition*	*SHIP Output Data files*
3	HID	Household unique ID	11_EXPHOTL.dta
E59	HOTCAT	Catering services, including eating out	11_EXPHOTL.dta
E60	HOTACC	Accommodation services	11_EXPHOTL.dta
E61	TOTHOTL	Total Restaurants and hotel expenditures, sum of the variables with HOT	11_EXPHOTL.dta

TABLE 4.12: MISCELLANEOUS GOODS AND SERVICES

All variables are annualized and in current local nominal prices

	Variable name	*Variable definition*	*SHIP Output Data files*
3	HID	Household unique ID	12_EXPMISC.dta
E62	MISPERSN	Personal care, e.g., hairdressing, electronic appliances for personal care.	12_EXPMISC.dta
E63	MISPROST	Spending on prostitution	12_EXPMISC.dta
E64	MISPEREFF	Personal effects not mentioned elsewhere	12_EXPMISC.dta
E65	MISSPROT	Contribution to social protection	12_EXPMISC.dta
E66	MISINSUR	Premium paid for insurances, including housing, health, etc.	12_EXPMISC.dta
E67	MISFINAN	Financial services	12_EXPMISC.dta
E68	MISOTHER	Other services not mentioned elsewhere. **Include Ceremonies such as weddings/funerals etc. if they are not found to be lumpy expenditures i.e. more than 5% of the total households incurred it.**	12_EXPMISC.dta
E69	TOTMISC	Total miscellaneous spending, sum of all variables with the prefix MIS	12_EXPMISC.dta

TABLE 4.13: REGIONAL PRICE DEFLATORS (OPTIONAL)

Normally regional food price and/or price deflators are included in the original survey data files. In this case, do not create FDINEX from the survey data. Just use the food price index/price index in the original survey data as FDINDEX. If there is no price deflator in the original survey and there are good price data in the survey, then FDINDEX can be created using tables 4.13.1 and 4.13.2. However, if there is no good price data in the survey, which is often the case, FDINDEX should be left missing and no regional deflated consumption will be available at the final expenditure file.

Table 4.13.1 Sub-File

	Variable name	***Variable definition***	***Syntax and data files***
A	REGION	ID for regions, the lowest level to calculate price index	13a_Pindex1.dta
B	FDID	ID for items, select 20 items that have the largest food share	13a_Pindex1.dta
C	FDRPRICE	Regional food prices for the 20 selected items	13a_Pindex1.dta
D	FDSHR	Normalize the 20 item food shares to 100 percent	13a_Pindex1.dta
E	FDINDEX	Regional food price deflator.	13a_Pindex1.dta

Note:

a) The FDINDEX will be MERGED into CCD_YEAR_E.dta based on the lowest level to calculate price index, normally at regional level. Therefore households in the same region will have identical values for FDINDEX.

b) FDINDEX is calculated as $FDINDEX = \frac{\sum_{i=1}^{20} pr^i \times q^i}{\sum_{i=1}^{20} p^i \times q^i}$ where pr^i is the regional price, p^i is national average price and q^i is the national average food share.

Table 4.13.2Food Price Index

	Variable name	***Variable definition***	***Data files***
2	REGION	ID for regions, the lowest level to calculate price index	13_FDPINDEX.dta
E70	FDPINDEX	Regional food price deflators.	13_FDPINDEX.dta

TABLE 4.14: TOTAL ANNUAL HOUSEHOLD EXPENDITURES (CCD_YEAR_E.DTA)

Merge Tables 1.0, 4.1 – 4.13

	SHIP Variable	***SHIP Variable definition***
1	COUNTRY	Country code
2	REGION	Region code and label the names
3	HID	Household unique identification
4a	SURVEMO	Month of survey
4b	SURVEYR	Year of survey
5	RURURB	Area of residence: *1 = Rural*; *2 = Urban*
6a	HHSIZE	Household members
6b	HHSIZE_S	Household members including servants
7	WTA_S	Individual weighting coefficient
8	WTA_S_HHSIZE	Household weighting coefficient: WTA_S_HHSIZE=WTA_S * HHSIZE
E6	TOTFOOD	Food and non-alcoholic beverages
E11	TOTALCH	Alcoholic beverages, tobacco and narcotics
E66	TOTFDAL	Sum of TOTFOOD(E6) and TOTALCH(E11)
E14	TOTCLTH	Clothing and footwear
E24	HSUTILITY	Electricity and gas
E19	HSFUEL	Firewood, and charcoal
E20	GAS	Gas for cooking
E21	ELEC	Electricity
E22	HSKEROSENE	Kerosene (Paraffin)
E23	HSDIESEL	Diesel for all purposes except transportation
E25	TOTHOUS	Housing, water, electricity, gas and other fuels
E32	TOTFURN	Furnishings, household equipment, and routine household maintenance
E36	TOTHLTH	Health
E38	TRFUEL	Fuel for transportation, such as petrol, diesel, etc…..
E 39	TRPERSON	Operation of personal transport equipment, such as gas, repairs and hire of chauffeurs
E 40	TRSERVE	Transport services: public transportation, purchase of airplane tickets, etc.
E41	TOTTRSP	Transportation
E45	TOTCMNQ	Communication
E52	TOTRCRE	Recreations and culture
E58	TOTEDUC	Education
E61	TOTHOTL	Restaurants and hotels
E69	TOTMISC	Miscellaneous goods and services
E71	TOTNFD	Sum of all non-food expenditure variables
E70	FDPINDEX	Regional food price deflators
E72	HHEXP_N	Total annual nominal household expenditures
E73	HHEXP_R	Total annual real household expenditures, TOTFOOD is deflated by FDPINDEX, and non-food expenditures are not deflated
E74	CPI2005_DEF	CPI to deflate/inflate the HHEXP_R to 2005 local prices
E75	PPP2005	2005 PPP exchange rate to 2005 local currency

Note: All missing values for expenditure aggregates must be recoded to 0 in the Total Annual Expenditure file (CDD_YEAR_E.dta)

CHAPTER 5: DOCUMENTATION AND COMPUTER PROGRAMS

For each harmonized data file, a short document should be written, containing at least the following information:

a) Date of production of the harmonized data files.
b) Person responsible for production.
c) Extrapolated population, by Urban/Rural area.
d) Definition of the variable HID (this variable is usually the concatenation of several variables in the original dataset; these variables must be clearly identified, in order to facilitate the link with the original dataset).
e) All data editing checks (cleaning and assumptions) should be documented in the computer programs as comments.
f) Imputations and assumptions made (number of cases and method) should be recorded meticulously.
g) Computer programs should be attached in sequence.
h) Comments and warnings to the users (limits in survey coverage, reliability of data, etc).
i) Any other information pertaining to the data that might be useful to data users.

APPENDIX I: THE METHODOLOGY OF CONSTRUCTING EMPLOYMENT VARIABLES IN SHIP

The construction of employment and labor participation variables is specific to Sub-Saharan African context since over 80 percent of employment activities are in the informal sector. Studying labor participation in Tanzania, Bardasi, Beegle, Dillon and Serneel found that due to poor questionnaire design many unpaid family workers under reported their economic activities, especially women who reported domestic duties as the main activity. These individuals inevitably undertake some unpaid economic activities such as cultivating, raising livestock, fetching water, and collecting wood (preparing meals for the family and caring for own children are not classified as economic activities by ILO definition.) The Survey-Based Harmonized Indicator Program (SHIP) developed complementary steps to capture these under-reported economic activities.

Additionally, due to informal economic activities and under-reported employment, there are often many missing values for the industry of employment based on ISIC code. To remedy this situation we create a variable to classify industry into farm and non-farm sectors that can be gleaned from other modules of the survey, such as farm, household enterprises and time use modules.

The steps outlined below are designed to capture, to the greatest extent possible, the actual employment status, including women who work from home and take care of household responsibilities at the same time, and students who help with fetching wood and water.

Construction of labor variables based on SHIP methodology

In the SHIP manual, we construct labor force participation (LABFORCE_WB) first (Table 1), followed by the variables EMPLOYED_WB (Table 2). Please note that because seasonality is a significant issue in capturing informal employment we code SHIP employment variables based on 12 month information, followed by 7day/two-week information. We follow ILO definition broadly with supplementary steps to capture under-reported employment. It is important to keep in mind that supplementary steps only replace missing values generated from previous steps.

The final employment variable of interest is EMP_CAT_1 (Table 5), which provides statistics to monitor structural changes in employment, classified into five categories as follows:

EMP_CAT_1 (Table 5)

11 Wage public
12 Wage private non-agriculture
13 Wage private agriculture
21 Self-employed non-agriculture (household enterprises)
22 Self-employed agriculture (farmers).

It should be noted that there is a small overlap between employment by SHIP definition and unemployment by ILO definition. By ILO definition, anyone who is without a job and looking for a job in a reference period (normally from 7 to 2 weeks) should be classified as unemployed. However, these individuals may have worked in the last 12 months, classified as employed by SHIP definition. Because of the large proportion of informal employment and the almost absence of unemployment benefits, the status between employed and unemployed is often blurred and has no practical implications. For this reason, this small overlap has no significance in analyzing labor force participation.

Below are summary tables on construction of most important SHIP employment variables. The sequence corresponds to the precedence of information, i.e. the later steps only replace missing values generated by previous steps.

Table 1: Construction of **LABFORCE_WB**

Variable Names			Description
LABFORCE_WB	**YES,** in the labor force, if **ANY** of the conditions is YES	WORKED_12=YES	Did the Individual have any kind of employment for any duration of time within the last 12 month recall period?
		WORKED_7=YES	Did the individual have any kind of employment for any duration of time within the last 7 day recall period?
		ABSENT=YES	If the individual did not work in the past 7 days did he/she have a job to return to?
		LOOKJOB=YES	Did the individual look for a job in the reference period (7days, 2 weeks or 4 weeks)?
		EMPTYPE_WB_1≠99	Type of employment based on information in the employment section and other sections of the survey. (Table 3 and Flowchart 1).
	NO	WORKED_12=NO WORKED_7=NO ABSENT=NO LOOKJOB=NO EMPTYPE_WB_1=99	If **ALL** the conditions are satisfied.
	MISSING	WORKED_12=. WORKED_7=. ABSENT=. LOOKJOB=. EMPTYPE_WB_1=99	If **ALL** the conditions are satisfied.

Table 2: Construction of **EMPLOYED_WB**

Variable Names			Description
EMPLOYED_WB	**YES,** if **ANY** of the conditions are satisfied	WORKED_12=YES	Did the Individual have any kind of employment for any duration of time within the last 12 month recall period? It is based on an employment screening question, otherwise missing.
		WORKED_7=YES	Did the individual have any kind of employment for any duration of time within the last 7 day recall period? It is based on an employment screening question, otherwise missing.
		ABSENT=YES	If the individual did not work in the past 7 days did he/she have a job to return to?
		EMPTYPE_WB_1≠99	Type of employment based on information in the employment section and other sections of the survey. (Table 3 and Flowchart 1).
	NO	LABFORCE_WB=YES	If the individual is part of the labor-force but does not have a job as identified by the above four variables.
	MISSING	LABFORCE_WB=.	If **no** information on labor force participation.

Description of EMPTYPE_WB_1

EMPTYPE_WB_1 is the most important variable in the SHIP Labor-force module. Information is sought form all sections of the survey including the household enterprise and the farming sections. By design it has no missing values. Individuals for whom no information can be found (include “not known,”, “other”, and no response/not applicable) are coded as 99. EMPTYPE_WB_1 classifies the type of employment into 8 different types.

Table 3: Description of codes in **EMPTYPE_WB_1**

Code	Label	Definition
1	Wage and salaried worker	An individual employed by non-household members who are paid in cash or in kind on a regular basis or based on a task. Agricultural and non-agricultural laborers are included in this category.
2	Self-employed with employees	An individual who runs a farm or a non-agricultural enterprise and employs at least one non-household member. Some surveys have information on employment of non-household members and is u8sed to define this category.
3	Self-employed without employees	An individual who runs a farm or a non-agricultural enterprise and DOES NOT employ any non-household member. If the survey does not have information on employment of non-household members then individuals are by default classified in this category. Some surveys have a category called 'own account worker'; such individuals are categorized in this section.
4	Employer	Refers to the owner of a business with employees, irrespective of agricultural or non-agricultural sector. Individuals are classified only if the survey explicitly has 'employer' as a category.
5	Domestic employees	An individual who works for a domestic household. Some surveys that have a question on job description may have this information.
6	Family worker	An individual who is a paid or unpaid worker who assists in the work on a farm or a non-farm enterprise.
7	Apprentice	Individuals who are apprentices, irrespective of whether paid or unpaid
9	Volunteer	An individual who is a volunteer, stated as an explicit category in the questionnaire.
99	Other/Not known/missing	All individuals who respond as other or not known or have a non-response value. Additional information from other sections of the survey is gleaned to reclassify these individuals into one of the above mentioned categories where-ever possible.

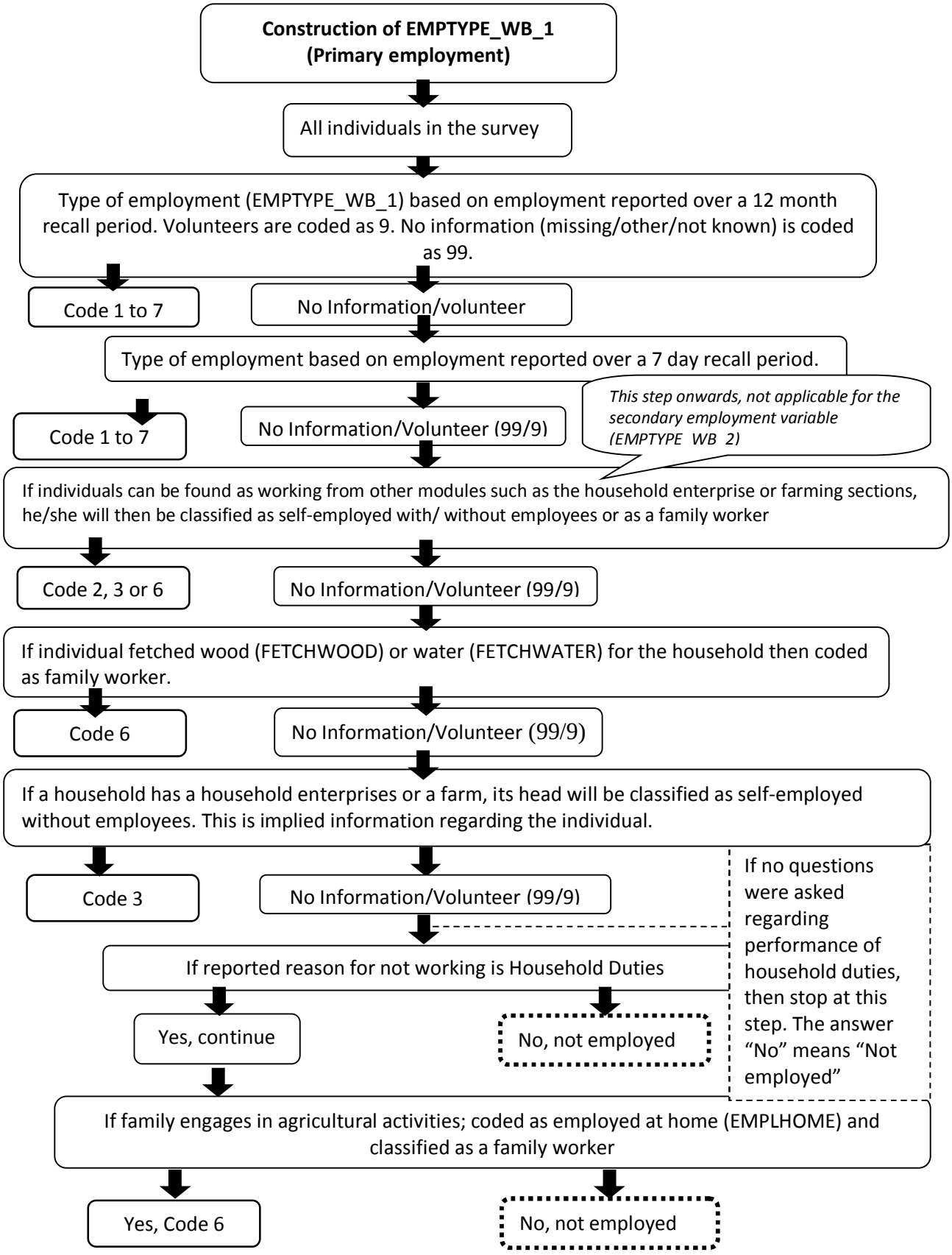

Construction of EMPTYPE_WB_1
(Primary employment)
All individuals in the survey
Type of employment (EMPTYPE_WB_1) based on employment reported over a 12 month recall period. Volunteers are coded as 9. No information (missing/other/not known) is coded as 99.
Code 1 to 7
No Information/volunteer
Type of employment based on employment reported over a 7 day recall period.
Code 1 to 7
No Information/Volunteer (99/9)
This step onwards, not applicable for the secondary employment variable (EMPTYPE_WB_2)
If individuals can be found as working from other modules such as the household enterprise or farming sections, he/she will then be classified as self-employed with/ without employees or as a family worker
Code 2, 3 or 6
No Information/Volunteer (99/9)
If individual fetched wood (FETCHWOOD) or water (FETCHWATER) for the household then coded as family worker.
Code 6
No Information/Volunteer (99/9)
If a household has a household enterprises or a farm, its head will be classified as self-employed without employees. This is implied information regarding the individual.
Code 3
No Information/Volunteer (99/9)
If no questions were asked regarding performance of household duties, then stop at this step. The answer "No" means "Not employed"
If reported reason for not working is Household Duties
Yes, continue
No, not employed
If family engages in agricultural activities; coded as employed at home (EMPLHOME) and classified as a family worker
Yes, Code 6
No, not employed

Table 4: Description of **EMPFRM** and **EMPSEC**

EMPFRM	Coded as 'yes' if the individual is employed in agricultural activities. Activities in the agriculture sector include regular farming, sharecropping, raising livestock, bee-keeping, fishing, logging and hunting. Individuals who fetch wood and/or water (FETCHWOOD/FETCHWATER) for the household or are employed at home(EMPLHOME), are classified in the agriculture sector. In addition, if the household engages in agricultural activities, then the heads of such households are classified to be in the agriculture sector.
EMPSEC	Captures the type of the employment establishment – 1) public, 2) private and 3) State-owned enterprise. It is based only on an explicit question in the survey to this effect.

EMP_CAT_1 is derived from the following underlining variables:

1. EMPTYPE_WB_1 (Table 3 and Flowchart 1)
2. EMPFARM_1 (Table 4)
3. EMPSEC_1 (Table 4)

Table 5: Construction of **EMP_CAT_1**

Variable Names			Description
EMP_CAT_1	EMP_CAT_1=11 Wage public	EMPTYPE_WB_1= 1, 4, 5, 7, 9	Type of employment is wage or employer or domestic employee or apprentice or volunteer.
		EMPSEC_1= 1 , 2	Individual worked in either the public sector or in a state owned company.
	EMP_CAT_1=12 Wage private non-agriculture	EMPTYPE_WB_1= 1, 4, 5, 7, 9	Type of employment is wage or employer or domestic employee or apprentice or volunteer.
		EMPSEC_1= 3	Individual worked in the private sector.
		EMPFRM_1≠YES	Individual was **NOT** employed in agricultural activities.
	EMP_CAT_1=13 Wage private agriculture	EMPTYPE_WB_1= 1, 4, 5, 7, 9	Type of employment is wage or employer or domestic employee or apprentice or volunteer.
		EMPSEC_1= 3	Individual worked in the private sector.
		EMPFRM_1=YES	Individual was employed in agricultural activities.
	EMP_CAT_1=13 Family enterprise non-agriculture	EMPTYPE_WB_1= 2, 3, 6	Type of employment is self-employed with or without employees or family worker.
		EMPFRM_1≠YES	Individual was **NOT** employed in agricultural activities.
	EMP_CAT_1=13 Family farmer	EMPTYPE_WB_1= 2, 3, 6	Type of employment is self-employed with or without employees or family worker.
		EMPFRM_1 =YES	Individual was employed in agricultural activities.

APPENDIX II: ISSUES IN HOUSEHOLD EXPENDITURE AGGREGATION

DATA CLEANING

Eliminating duplicates: Before the aggregation, one should make sure that there is no duplicated household or individual id (This problem should have been eliminated at the data entering stage, this is just a quick check). There are two types of duplicate household id. One is that one household was entered twice, in which case all other variables should be also the same, such as physical features of the housing, rent paid, etc. In such case, one of the duplicated observations should be eliminated (STATA) can perform this task easily). Second type of duplicate is two or more households have identical households ID. This type of duplicate is hard to discover because one often takes two households as one household. However, unusual large household size, such as over 30, should warrantee a check of the household to make sure that they are not two households. When there are two or more same ids signed to household members, if the two or more members have exact same information, such as age, sex, status enrollment, etc., a double entry of the same member is usually the case, and one should eliminate one of the observations. However, if they are two different individuals, one should reassign a new id to one of the members.

Cleaning expenditure per capita variable: Per capital annual food and non-food consumption is first adjusted by regional food price indices. Outliers (separately for food and non-food expenditure), greater than three standard deviations will then be replaced with the median values of the corresponding per capita food and non-food consumption, by region and by rural/urban areas. Total number of observations cleaned should be less than 1 percent of total observations. If it is greater than 1 percent of total observations, aggregation procedures should be rechecked to ascertain accuracy.

Do not assign missing values: A common mistake is to assign missing values with some meaningful values when aggregating a variable. To avoid incorrectly assigning missing values, avoid open ended grouping while coding. For example, do not use "if AGE<5 then CHDUNDR5=1", instead, use "if 0<=AGE<=5 then CHDUNDR5=1". In addition, the number of missing values must be checked before and after the aggregation to see if the numbers of missing values are consistent.

DATA AGGREGATION

Purchased Food expenditure aggregation in general is straight forward. For purchased food, one just adds up all values of relevant items according to the sub-categories. If the recall period is 2 weeks, the annualized expenditure will be derived by multiplying the reported values by 365/14, and if the recall period is one month, then by 365/30.

Auto food expenditure aggregation is more complicated. Food auto-consumption comes in two varieties: that which is produced by the household and the gifts in kind which the household consumes. Both must be valued, although the first may be the only one to immediately spring to the reader's mind.

If the total value of auto-consumption is not asked, in general, the valuation of auto-consumption is a two-step process. First, the prices are imputed for the goods that are auto-consumed, and then the aggregation is done as for purchased expenditure. If the survey had a community price questionnaire, this is usually the source of prices for auto-consumption. If such information is missing but prices can be inferred from household survey on auto consumption, then this is the source for prices. If neither method is possible, external sources of price data are used. An example is administrative records of market prices. In all these cases, there is some degree of bias, since the market price for a good is not necessarily farm gate prices. There may be an intermediary who purchases at less than the market

price, or there are travel and time costs associated with selling the good at market. However, it is better to have the information and (potentially) over-estimate the value than to ignore it entirely.

Generally, surveys collect information on food expenditure using multiple units of measure. A common unit of measure is necessary in order to compare prices; the simplest are often metric units like kilos and liters. Note that prices will not be the same for all units of measure: buying rice by the gram is likely to be more expensive than buying it in 50-kg bags (volume discount).

Once the price data is found, the level of disaggregation at which the prices are useful must be determined. The tradeoff is one of number of observations versus the household's reality. As the number of observations increases, the confidence in the mean (or median) price increases too; hence the imputed price is better. However, all enumeration areas in a district need not have the same market and growing conditions; the less fine the geography, the greater the risk of "noise" due to dissimilarities. This is a fine line, but generally 30 observations in for each product in each region are a minimum.

Once the prices are determined, the remainder of the work is identical to that of the purchased food consumption, except that the imputed prices are used to value the consumption. For purposes of validation and analysis, it is recommended that the same categories as for purchased expenditure be used.

Education expenditure aggregation should be annualized based on the time of school year. For example, monthly transportation fee should only be multiplied by the number of the months when children go to school, normally 9, to get the annual transportation expenditure.

Health care expenditure aggregation is more difficult. Health care expenditures are often asked for the last three months or so. Although if one was sick in the last three months, it does not necessarily mean that one was sick every three months. In practice, however, the three-month expenditure is often multiplied by 4 to get the annualized expenditure. This is more or less fine with expenditures on consultation fees and medication. For hospitalization fees, however, if it is lumpy, it should be treated as a onetime annual expenditure.

Housing expenditure is often only available for households who rent. For owner occupied houses, if mortgage or the market value of the rent is not asked code it as missing. [3]

Durable goods and Lumpy expenditures are excluded: This rule needs to be implemented with caution: in some countries one may find that many households report weddings, baptisms, and funerals in a survey. In the case of the 1994-95 Enquête Permanente de Conjoncture Economique et Sociale in Niger, over 60 percent (weighted) of all regions' households reported expenditure related to a marriage or a baptism. In this case, it is no longer an infrequent, uncommon expenditure item. The general rule

[3] The most commonly used methodology to impute rent is to use a hedonic regression to estimate the housing value for owner occupied houses. First, a regression model is fitted to households who paid rent, with dependent variable being the reported monthly rent, and independent variables being the physical features of the households, including latrines, number of rooms, materials of floor, wall and roof, the presence of a kitchen, the source of drinking water, the source of lighting power, and the cooking fuel. Other variables, known as "controls" are added to the regression to adjust for regional or rural-urban differences.

Once the model is done, a useful check of the quality is to use it to predict the rent for households that paid rent. If there is a systematic difference, something remains to be exploited. Otherwise, the desired result is small deviations between the actual and predicted rents. The parameters estimated from the hedonic regression are then used to estimate the rent for owner occupied houses. For the final rent that will be included in the total household expenditure, there are two choices. One is to use rent predicted by the regression for all the households, including those who paid rent. The reason of doing so is that regression often predicts rent somewhat lower than the actual rent paid. Using regression estimated rent for all the households puts everyone on equal footing. Another one is to use regression estimated rent for the owner occupied households only, the actual rent paid is used for renters.

is to omit lumpy expenditure, the ones that are not frequently or commonly spent, especially if they are large relative to total household expenditure, from the total household expenditure aggregate.

Adjustment of cost of living is to take into account that same money can buy different amount of goods in different regions. For example, a pound of beef would be much more expensive in a capital city than in a rural town where the cattle are raised nearby. To calculate regional price deflators, national average price and food share for the largest 20 items will be used as base price index. The regional price deflator is then calculated as where is price deflator for region r, is the national share of commodity "k" for the national food basket, P_r^k if the regional price for commodity "k", and is the national average price for commodity "k".

Common Sense: Expenditure aggregation is easier said than done, many judgments must be made on individual cases. The most important rule perhaps is to use on the ground knowledge/ common sense while using statistical tools. The second important rule is to document in detail the steps used for aggregation. Finally, some basic statistics should be checked against reality.

ANNEX III: ISO 3166-1 ALPHA-3 - COUNTRY CODES (SUB-SAHARAN AFRICA)

AGO	Angola	MDG	Madagascar
BEN	Benin	MWI	Malawi
BWA	Botswana	MLI	Mali
BFA	Burkina Faso	MRT	Mauritania
BDI	Burundi	MUS	Mauritius
CMR	Cameroon	MOZ	Mozambique
CPV	Cape Verde	NAM	Namibia
CAF	Central African Republic	NER	Niger
TCD	Chad	NGA	Nigeria
COM	Comoros	RWA	Rwanda
COD	Congo, Dem. Rep.	STP	Sao Tome and Principe
COG	Congo, Rep.	SEN	Senegal
CIV	Cote d'Ivoire	SYC	Seychelles
GNQ	Equatorial Guinea	SLE	Sierra Leone
ERI	Eritrea	SOM	Somalia
ETH	Ethiopia	ZAF	South Africa
GAB	Gabon	SSD	South Sudan
GMB	Gambia, The	SDN	Sudan
GHA	Ghana	SWZ	Swaziland
GIN	Guinea	TZA	Tanzania
GNB	Guinea-Bissau	TGO	Togo
KEN	Kenya	UGA	Uganda
LSO	Lesotho	ZMB	Zambia
LBR	Liberia	ZWE	Zimbabwe

ANNEX IV: INTERNATIONAL STANDARD INDUSTRIAL CLASSIFICATION (ISIC)

The following is an extract from the International Standard Industrial Classification of Economic Activities[4] (ISIC) Rev. 3.1. It will be used to generate variables INDUSTRY_1 and INDUSTRY_2.

The individual categories of ISIC have been aggregated into the following 17 sections:

Section	Divisions	Description
A	01, 02	Agriculture, hunting and forestry
B	05	Fishing
C	10-14	Mining and quarrying
D	15-37	Manufacturing
E	40, 41	Electricity, gas and water supply
F	45	Construction
G	50-52	Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods
H	55	Hotels and restaurants
I	60-64	Transport, storage and communications
J	65-67	Financial intermediation
K	70-74	Real estate, renting and business activities
L	75	Public administration and defense; compulsory social security
M	80	Education
N	85	Health and social work
O	90-93	Other community, social and personal service activities
P	95-97	Activities of private households as employers and undifferentiated production activities of private households
Q	99	Extraterritorial organizations and bodies

The Classifications registry keeps updated information on Statistical Classifications maintained by the United Nations Statistics Division (UNSD). For the SHIP we will use ISIC rev. 3.1. For generating EMP_IND, the classification is as following:

For standardized file, this classification is used:-		
A + B	=	Agriculture and fishing
C	=	Mining and quarrying
D	=	Manufacturing
E	=	Electricity, gas and water supply
F	=	Construction
G + H	=	Commerce
I	=	Transport, storage and communication
J + K	=	Financial, insurance and real estate
L	=	Public administration
M, N	=	Education, health and social work
O + P + Q	=	Other services

[4] http://esa.un.org/unsd/cr/registry/regcst.asp?Cl=17&Lg=1

REFERENCES

Ahmad S., (1980), *Approaches to purchasing power parity and real product comparisons using shortcuts and reduced information*, World Bank Staff working paper No. SWP 418 1980/10/01

Berghahl M., Japec L., Madaleno M., Signore M., Tzougas I., *Tools for Standardising and Improving the Quality of Statistics Production Processes and other Operations,* Paper presented during the Stockholm 2001 conference, May 14-15, 2001.

Carson C. S., (2001), *Toward a Framework for Assessing Data Quality*, Paper presented during the Stockholm 2001 conference, May 14-15, 2001.

Deaton A. and Zaidi S., (2002), Guidelines for Constructing Consumption Aggregates for Welfare Analysis, World Bank *LSMS Working Paper No. 135*

Dupriez O., (2005), Poverty PPPs: Building a Database on Household Consumption Profiles by ICP Basic Heading.

Eurostat, *Assessment of the Quality in Statistics*, Paper presented during the Stockholm 2001 conference, May 14-15, 2001.

Haworth M., Bergall M., Booleman M., Jones T. and Madaleno M., (2001), *Leadership Expert Group on Quality; Quality Framework Chapter*, Paper presented during the Stockholm 2001 conference, May 14-15, 2001.

http://www.cdc.gov

http://esa.un.org/unsd/cr/registry/regcst.asp?Cl=17&Lg=1

http://unescostat.unesco.org/en/stats/stats0.htm

http://www.ilo.org/public/english/bureau/stat/class/icse.htm

http://www.ilo.org/public/english/bureau/stat/res/ecacpop.htm

http://www.ilo.org/public/english/employment/skills/informal/index.htm

http://www.who.int/nutgrowthdb/reference/en/

Hussmanns R., Mehran F., Verma V., (1990), *Surveys of economically active population, employment, unemployment and underemployment*: An ILO manual on concepts and methods.

Irving Kravis B., (1986), The three faces of the International Comparison Project, *Research Observer* 1 No. 1 January 1986.

Isenman P., (1979), *Inter-country comparison of "Real" (PPP) incomes : revised estimates and unresolved questions*, World Bank Staff working paper No. SWP 358 1979/09/01.

Japec L., *Survey of Quality Practices in National Statistical Institutes*, Paper presented during the Stockholm 2001 conference, May 14-15, 2001.

Kravis I. B., (1986), The Three faces of the International Comparison Project, *Research Observer 1, no. 1*.

Lanjouw P., Prennushi G., Zaidi S. (1996), Building Blocks for a Consumption-Based Analysis of Poverty in Nepal, Washington DC, World Bank.

Litchfield J. A., (1999), Inequality: methods and Tools, World Bank.

Minhas, B. S., (1988), "Validation of large-scale sample survey data: case of NSS household consumption expenditure," Sankhya, *Series B 50, Supplement, 1-63*.

Mungai R. and Marchant T., (2003), Establishing Comparability of Household Data across Countries and Time, Topic IPM 35 Designing Surveys to Provide Cross-National Comparable Data, *Bulletin of the International Statistical Institute 54th Session*.

Ravallion M., (1998), Poverty Lines in Theory and Practice, World Bank *LSMS Working Paper No. 88*

Ravallion M., (1992), Poverty Comparisons: A Guide to Concepts and Methods, World Bank *LSMS Working Paper No. 133*

Statistics Canada, (1998), Third Ed., *Statistics Canada Quality Guidelines*,

World Bank, (2007), *World Development Indicators 2007*, Washington DC, World Bank.

World Bank, (2001), *World Development Indicators 2001*, Washington DC, World Bank.

World Bank, (1999), Poverty Reduction and the World Bank: Progress in Fiscal 1998, Washington DC, World Bank.S

APPENDIX V: STANDARD EXCEL SPREADSHEET, *to be continued*

Country Survey Name Year: Selected Social & MDG Indicators by Rural Urban and Consumption Level																		
		Page 1 of 2																
				National			**Rural**						Urban					
Country name	Year	***Indicators***	Unit of measure	**Poorest 20%**	Richest 20%	**All**	**Poorest 20%**	II	III	IV	Richest 20%	**All**	**Poorest 20%**	II	III	IV	Richest 20%	**All**
Country name	Year	**101.*Demographic indicators**																
Country name	Year	102. Sample size (households)	Number															
Country name	Year	103. Total population (weighted)	Million															
Country name	Year	104. Age dependency ratio (age <15 or >64)	Percent															
Country name	Year	105. Average household size	Number															
Country name	Year	**106. *Head of household characteristics**																
Country name	Year	**107. *Marital status of head**																
Country name	Year	108. Married/common law male	Percent															
Country name	Year	109. Single/ divorced/widowed male	Percent															
Country name	Year	110. Married/common law female	Percent															
Country name	Year	111. Single/ divorced/widowed female	Percent															
Country name	Year	**112. *Education level of head**																
Country name	Year	113. No level	Percent															
Country name	Year	114. Preschool/Primary not completed	Percent															
Country name	Year	115. Primary completed	Percent															
Country name	Year	116. Lower secondary completed	Percent															
Country name	Year	117. Senior secondary completed	Percent															
Country name	Year	118. Post secondary technical																
Country name	Year	119. Tertiary	Percent															
Country name	Year	120. Other	Percent															
Country name	Year	121. Missing	Percent															
Country name	Year	**122. *Labor force**																
Country name	Year	**123. *Employment (15 years and above)**																
Country name	Year	**124. Proportion of total population aged 15 yrs or more**	Percent															
Country name	Year	125. Employed	Percent															
Country name	*Year*	*126. of which females*	*Percent*															
Country name	Year	**127. *Employment Categories (15 and above) Primary employment**																
Country name	Year	**128. *Wage workers**																
Country name	Year	129. Wage public	Percent															
Country name	*Year*	*130.of which females*	*Percent*															
Country name	Year	131. Wage private non-agriculture	Percent															
Country name	*Year*	*132. of which females*	*Percent*															
Country name	Year	133. Wage private agriculture	Percent															
Country name	*Year*	*134. of which females*	*Percent*															
Country name	Year	**135. *Self employed/ non-wage**																
Country name	Year	136. Family enterprise	Percent															
Country name	*Year*	*137. of which females*	*Percent*															
Country name	Year	138. Family farmer	Percent															
Country name	*Year*	*139. of which females*	*Percent*															
Country name	Year	**140. *Household Sources of Income (Primary & Secondary emplymt)**																
Country name	Year	141. Agriculture	Percent															
Country name	Year	142. Non-Agriculture non-wage	Percent															
Country name	Year	143. Wage	Percent															

Standard Excel Spreadsheet, *continued*

Country name	Year	**144. *MDG1: Eradicate extreme poverty and hunger**	
Country name	Year	145. Mean monthly per capita expenditure	PPP2005
Country name	Year	146. Poverty headcount under $PPP1.25/day	Percent
Country name	Year	147. Gini coefficient, consumption	Ratio
Country name	Year	**148. *Mean share of household expenditures**	
Country name	Year	149. Food expenditure	Percent
Country name	Year	150. Health expenditure	Percent
Country name	Year	151. Education expenditure	Percent
Country name	Year	152. Public transportation expenditure	Percent
Country name	Year	153. Private transportation expenditure (excludes fuel)	Percent
Country name	Year	154. Fuel expenditure (for transportation)	Percent
Country name	Year	155. Electricity expenditure	Percent
Country name	Year	156. Gas expenditure	Percent
Country name	Year	157. Kerosene expenditure	Percent
Country name	Year	158. Diesel expenditure (all purposes but transport)	Percent
Country name	Year	**159. *MDG2: Education and Literacy; MDG3 Promote gender equality**	
Country name	Year	**160. * Gross primary enrollment rate**	
Country name	Year	161. Female	Percent
Country name	Year	162. Male	Percent
Country name	Year	**163. *Gross lower secondary enrollment ratio**	
Country name	Year	164. Female	Percent
Country name	Year	165. Male	Percent
Country name	Year	**166. *Gross senior secondary enrollment ratio**	
Country name	Year	167. Female	Percent
Country name	Year	168. Male	Percent
Country name	Year	**169*. Net primary enrollment rate**	
Country name	Year	170. Female	Percent
Country name	Year	171. Male	Percent
Country name	Year	**172. *Primary completion rate on time**	
Country name	Year	173. Female	Percent
Country name	Year	174. Male	Percent
Country name	Year	**175. *Net lower secondary enrollment ratio**	
Country name	Year	176. Female	Percent
Country name	Year	177. Male	Percent
Country name	Year	**178. *Net senior secondary enrollment ratio**	
Country name	Year	179. Female	Percent
Country name	Year	180. Male	Percent
Country name	Year	**181. *Tertiary enrollment over age 18-22***	Percent
Country name	Year	*182. of which female*	*Percent*
Country name	Year	**183. *Youth literacy rate ages 15-24**	
Country name	Year	184. Female	Percent
Country name	Year	185. Male	Percent
Country name	Year	**186. *MDG4: Reduce child mortality; MDG5: Improve maternal health**	
Country name	Year	187. Distance to health centers < 5 km	Percent
Country name	Year	188. Percent claiming to be sick last 2 wks	Percent
Country name	Year	*189. of which percent sought medical care*	*Percent*
Country name	Year	190. Birth assisted by trained staff	Percent
Country name	Year	191. 1-year-olds total immunization	Percent
Country name	Year	192. 1-year-olds immunization, measles	Percent
Country name	Year	**193. *MDG7: Ensure environment sustainability**	
Country name	Year	194. Owner occupancy rate	Percent
Country name	Year	195. Access to flush/improved toilets	Percent
Country name	Year	*196. of which improved pit lartrine*	*Percent*
Country name	Year	197. Access to garbage collection	Percent
Country name	Year	198. Connected to electricity	Percent
Country name	Year	*199. of which solar/wind electricity*	*Percent*
Country name	Year	**200. *Fuel use**	
Country name	Year	201. Firewood/charcoal	Percent
Country name	Year	202. Kerosene/electricity/gas	Percent
Country name	Year	203. Other	Percent
Country name	Year	204. Missing	Percent
Country name	Year	**205. *Access to water**	
Country name	Year	206. Improved	Percent
Country name	Year	207. Unimproved	Percent
Country name	Year	208. Other	Percent
Country name	Year	209. Missing	Percent