Documentation LSMS-ISA Malawi 2019

Romain Frelat

27-05-2024

Objectives

This document records the steps and choices made to transform LSMS-ISA Malawi 2019 dataset into the format of the Household Data Platform. It goes together and complements the script file `Get_LSMS_MWI2019.R`. The original dataset (MWI_2019_IHS-V_v05_M, DOI 10.48529/yqn3-zv74) can be accessed from the World Bank microdata library: https://microdata.worldbank.org/index.php/catalog/3818/ [last accessed 30/11/2023]

Crop table

The crop table contains 8 columns with information on crop cultivation and usage.

Name	Definition	Unit
hhid	household id	
name	name of the crop	
land_area_ha	land cultivated	hectare
harvest_kg	amount harvested	kg
consumed_kg	amount consumed	kg
sold_kg	amount sold	kg
income_lcu	income from sells	lcu
use_list	list all usages	categories

The crop information is divided in three different sections: the rainy season (module C and G), the dry season (module K and M), and the perenial crops (module P). The same transformations were made for the three sources of information, that were merged together after processing.

Land area

The field size were transformed into hectare. Most were measured by GPS and reported in *acre* ($1 acre = 0.404686 ha$). When not available, we used the household estimates and the following conversion factors.

```
##            ACRE         HECTARE   SQUARE METERS           YARDS OTHER (SPECIFY)
##    4.046860e-01    1.000000e+00    1.000000e-04    8.361204e-05              NA
```

Additionally, we used the reported percentage of the field cover by each crop (given in 5 categories) to estimate the area per crop. We corrected these percentages when necessary to make sure that the sum of the crop areas per field is not higher than the field area.

Harvest and uses

The crop harvest and usage are estimated from the post harvest questionnaire, in section 11. The conversion factors for units of weight are calculated from the file `ihs_seasonalcropconversion_factor_2020.csv`,

`ihs_foodconversion_factor_2020.csv`, and `ihs_treeconversion_factor_2020.csv`. We simplified and used only one conversion factor per unit (the median), without considering the different crop.

The amount of crop consumed was not reported in the survey for the dry season and the perenial crop. In these caases, we calculated it as the difference between the quantity harvested and the reported quantity used (sum of the quantities given as gift, as payment of debts, to feed livestock, for by-product fabrication, for seed production and lost).

The quantities are checked for consistency. When the amount consumed and sold was higher than the amount harvested, we used the sum of the amount consumed and sold as quantity harvested, unless the resulting yield was higher than 20t/ha. In such case (abnormally high quantities sold and consumed), we lower the quantities sold and consumed using the same proportions as originally reported but calculated on the amount harvested.

Livestock table

The livestock table contains 3 columns with information on livestock herd

Name	Definition	Unit
hhid	household id	
name	name of the livestock	
n	number of livestock kept	

The information on the herd structure is provided in section 8 of livestock questionnaire.

We grouped *local-hen*, *local-cock*, *chicken-layer/chicken-broiler* and *chicks* into **poultry**.
We grouped *cow*, *bull*, *ox*, *calf*, *steer/heifer*, and *improved* into **cattle**.
We grouped *hare*, *hare and mbira*, and *mbila* into **rabbit**.

Livestock production table

The livestock production table contains 7 columns with information on livestock productions:

Name	Definition	Unit
hhid	household id	
name	name of the livestock	
prod	livestock production	
harvest_kg	amount harvested	kg
consumed_kg	amount consumed	kg
sold_kg	amount sold	kg
income_lcu	income from sells	lcu

Milk, egg and manure

The livestock production is recorded in module S, recording chicken eggs, guinea fowl eggs, cow milk, manure, meat, and skins and hides production. For meat, manure, and skin and hides, we attributed the production to the livestock category with the largest herd (converted in TLU number).

```
##                 id              name   prod
## 1     CHICKEN EGGS           POULTRY    egg
## 2         COW MILK            CATTLE   milk
## 3 GUINEA FOWL EGGS TURKEY/GUINEA FOWL    egg
## 4           MANURE              <NA> manure
## 5             MEAT              <NA>   meat
## 6  SKINS AND HIDES              <NA>  other
```

Whole animals

The selling of whole animals (reported in module R1) is also considered as livestock product. The whole animals are always sold (no quantities consumed, the quantity harvested equals the quantity sold). We converted them into kg using the TLU conversion factors (1TLU=250kg).

```
##         BEEKEEPING             CATTLE  DONKEY/MULE/HORSE        DOVE/PIGEON
##                0.0              175.0              175.0                2.5
##               DUCK               GOAT                PIG            POULTRY
##                2.5               25.0               75.0                2.5
##             RABBIT               RATS              SHEEP TURKEY/GUINEA FOWL
##                2.5                2.5               25.0               12.5
```

Household information table

The initial household information table contains information on household composition, off farm activities, and food security.

Name	Definition	Unit
hhid	household id	
country	country of the survey	
year	year of the survey	
gps_lat	latitude in decimal degrees	°N
gps_lon	longitude in decimal degrees	°E
hh_size_members	size of the household in number of persons	
hh_size_mae	size of the household in male adult equivalent	MAE
head_age	age of the household head	
head_gender	gender of the household head	'f' or 'm'
off_farm_lcu	off farm income per year	lcu
off_farm_div	diversity of off farm activities	
hdds_score	household diet diversity score based on 10 groups	
foodshortage_count	number of months with food shortage	
foodshortage_months	name of the months with food shortage	
currency_conversion	conversion from local currency to power parity purchase usd	lcu/usd

General information

`hhid`is made of the prefix "LSMS_MWI_2019" and the variable `case_id`. This variable is a unique identifier for the household that can connect all tables (crop, livestock, and livestock production).

The GPS coordinates are provided with a rounding at 0.1 decimal degree. The currency power parity purchase conversion factor is provided by the World Bank for Ethiopia in 2018 (= 275.43mwk/usd).

Household size

The household composition is provided in section 1 of the household questionnaire. The male adult equivalent are calculated with 5 age-groups: 0-4, 5-10, 11-24, 25-50, 51+ and with the corresponding coefficient per gender and age class.

```
##        0-4 5-10 11-24 25-50  51+
## Male   0.5 0.75 0.925  1.00 0.73
## Female 0.5 0.75 0.750  0.86 0.60
```

Off farm income

We considered the employment (reported in section 4 of the household questionnaire) or businesses (reported in section 12). For employment, we calculated the number of hours worked, and the salary per hour based on the last payment. If the salary was higher than $1000 birr/h$ we set it to $1000 birr/h$. Then we multiplied the number of hours worked by the salary per hour to get the annual salary. If the annual salary resulted to be lower than the reported payment, we reported the payment as annual salary.

Food security

The diet diversity was calculated from the 7-day recall of food consumption (module G).

The months with food shortage was directly retrieve from questions in module H.

Crop and Livestock summary

The household information table also contains 28 columns with summary information from crop and livestock tables.

Name	Definition	Unit
hhid	household id	
land_cultivated_ha	total land cultivated	ha
crop_div	number of crop cultivated	
crop_name	names of crop cultivated	
crop_harvest_kg	total crop harvest	kg
crop_yield_kg_per_ha	crop yield	kg/ha
crop_sold_kg	quantity of crop sold	kg
crop_sold_perc	percentage of quantities of crop sold	%
crop_income_div	number of different crop sold	
crop_income_lcu	total income from crop production	lcu
crop_value_lcu	value of crop produced but not sold	lcu
crop_consumed_kcal	energy value from crop consumed	kcal
livestock_tlu	herd size	tlu
lstk_div	number of livestock species herded	
lstk_name	names of livestock species herded	
lstk_harvest_kg	total livestock product harvested	kg
lstk_sold_kg	quantity of livestock product sold	kg
lstk_sold_perc	percentage of livestock production sold	%
lstk_income_div	number of different livestock products sold	
lstk_income_lcu	total income from livestock production	lcu
lstk_value_lcu	value of livestock production not sold	lcu
lstk_consumed_kcal	energy value from livestock consumed	kcal
farm_div	number of crop and livestock species	
farm_harvest_kg	total farm production	kg
farm_sold_perc_kg	percentage of farm production sold	%
farm_income_div	number of different farm products sold	
farm_income_lcu	total income from farm production	lcu
tot_income_lcu	total income (farm + off farm)	lcu
farm_consumed_kcal	energy value from farm production consumed	kcal

All the values are calculated automatically with the function `calc_farm_prod()` from the package `farmhousehold`. The calculations are simple and summarize per households the quantities reported in the crop and livestock tables.

For energy conversion, we used estimates mostly from the FoodData Central of the U.S.Departement of Agriculture (https://fdc.nal.usda.gov/). Below are the energy conversion factors:

```
##           banana            beans          cabbage          cassava
##              890             1480              250             1600
##       groundnut            maize            mango         nkhwani
##             5500             3650              800              170
##            onion pigeonpeanandolo             rice         sorghum
##              720              810             3600             3390
##       soya_bean        sugarcane     sweet_potato        tanaposi
##             1470              400              770              150
##           tomato
##              210
```

GIS information

Based on the GPS coordinates of households, we extracted:

- the administrative region and district (from GADM)
- the Dixon farming system classification for Sub-Saharan Africa Dixon et al. 2021
- the population density estimated by the Gridded Population of the World (GPWv4)
- the travel time to cities estimated by Nelson et al. 2019
- the Koeppen's Climate Classification from:
  Beck, H.E., et al. (2018) "Present and future Köppen-Geiger climate classification maps at 1-km resolution", *Nature Scientific Data*, 5, 180214 DOI 10.1038/sdata.2018.214

Summary

We kept only *rural* households with at least one crop or one livestock. In total, the dataset is made of 9624 households, with information on 36501 crop cultivated and 8644 livestock species herd. The dataset made of the four tables, together with tlu and energy conversion factors are binded together into a `farmhousehold` object and saved into the file `HHDB_LSMS_MWI2019.rds`.