

README: Replication files for “Multidimensional Poverty: Why not Make Up the Missing Joint Distribution Data?”

Benoit Decerf, Mery Ferrando and Balint Menyhart.

March 15, 2026

Overview:

The code in this replication package constructs all tables and figures in the text and Online Appendix of the paper using Stata. A main.do file in the root directory serves as the single-entry point and automatically calls the two master do-files in sequence: 1) to clean the data (which calls 6 do-files to clean each of the household surveys used), and 2) to implement the quantitative analysis (which generates all the figures and tables and calls 2 other do-files for the simulation exercise).

Data Availability:

This study uses anonymized and nationally representative household survey data from six countries. All datasets are publicly available either directly from relevant national statistical authorities or the World Bank Microdata Library, subject to free registration on the respective data portals. Below we describe each dataset and provide access information.

1. Bolivia: Encuesta de Hogares (EH) 2015

Description: The Encuesta de Hogares 2015 is a multi-topic household survey conducted by the Instituto Nacional de Estadística (INE) of Bolivia. The survey collects information on housing characteristics, demographic composition, education, health, employment, income, and household expenditures, providing key indicators for monitoring poverty and living conditions. The sample contains 10,170 households across 9 regions.

Access policy: Publicly available (registration required)

Date accessed: 13/09/2019

Data URL:

<https://anda.ine.gob.bo/index.php/catalog>

2. Brazil: Pesquisa Nacional por Amostra de Domicílios (PNAD) 2015

Description: The PNAD 2015 is a nationally representative household sample survey conducted annually by the Instituto Brasileiro de Geografia e Estatística (IBGE). The survey covers general population characteristics, education, labor, income, and housing through 117,939 households across 27 provinces. The 2015 wave was the final edition of the traditional PNAD before its replacement by the continuous PNAD (PNAD Contínua).

Access policy: Publicly available (no registration required)

Date accessed: 13/09/2019

Data URL:

<https://www.ibge.gov.br/estatisticas/sociais/trabalho/9127-pesquisa-nacional-por-amostra-de-domicilios.html>

3. Ecuador: Encuesta de Condiciones de Vida (ECV) 2013–2014

Description: The ECV 2013–2014 is the sixth round of Ecuador's Living Conditions Survey, conducted by the Instituto Nacional de Estadística y Censos (INEC). The survey covers 28,970 households across all 24 provinces and collects information on housing, demographics, education, migration, employment, income, expenditures, health, social capital, time use, and subjective well-being.

Access policy: Publicly available (registration required)

Date accessed: 30/10/2019

Data URL:

<https://www.ecuadorencifras.gob.ec/encuesta-de-condiciones-de-vida-ecv/>

<https://www.ecuadorencifras.gob.ec/condiciones-de-vida-ecv-bases-de-datos/>

4. Ethiopia: Socioeconomic Survey (ESS) Wave 3, 2015–2016

Description: The Ethiopia Socioeconomic Survey (ESS) Wave 3 is a multi-topic panel household survey conducted by the Central Statistical Agency (CSA) of Ethiopia in collaboration with the World Bank Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA). The survey collects information on demographics, education, health, labor, welfare, agriculture, food security, and shocks from a nationally representative sample of 4,955 households.

Access policy: Publicly available (registration required)

Version: Version 03 (March 2020)

Reference ID: ETH_2015_ESS_v03_M

Date accessed: 09/08/2019

Data URL:

<https://microdata.worldbank.org/index.php/catalog/2783>

5. Ghana: Living Standards Survey Round 6 (GLSS6) 2012–2013

Description: The GLSS6 is the sixth round of Ghana's Living Standards Survey, conducted by the Ghana Statistical Service (GSS) from October 2012 to October 2013. The survey covers 16,772 households across 1,200 enumeration areas and collects detailed information on demographics, education, health, employment, migration, housing, agriculture, household expenditure, income, financial services, and governance.

Access policy: Publicly available (registration required)

Version: Version 1.0

Reference ID: GHA-GSS-GLSS6-2012-v1.0

Date accessed: 09/08/2019

Data URL:

<https://www2.statsghana.gov.gh/nada/index.php/catalog/72>

6. Uganda: National Panel Survey (UNPS) Wave 5, 2015–2016

Description: The UNPS 2015–2016 is the fifth wave of Uganda's multi-topic panel household survey, conducted by the Uganda Bureau of Statistics (UBOS) with support from the World Bank LSMS-ISA program. The survey is nationally representative and collects information on demographics, education, health, employment, income, consumption expenditure, agriculture, and community characteristics over a 12-month period with two household visits. The sample contains 3,314 households from 4 regional units.

Access policy: Publicly available (registration required)

Version: v02_M

Reference ID: UGA_2015_UNPS_v02_M

Date accessed: 31/08/2019

Data URL:

<https://microdata.worldbank.org/index.php/catalog/3460>

Data sources:

The main data files associated with each national survey are stored directly in the ``project'/data/in/` folder. Each survey has one main data file, as shown in the table below. (For more information on each of the surveys used, see the “data\_overview.xlsx” Excel-file saved directly under ``project'/data/`.) These files are the merged aggregate of all the relevant sub-components of the data distribution packages that are required for the analysis (See more information on how these are compiled below under the replication procedure heading.)

Dataset	Main file name:	Location:	Description (Country – Survey name – Reference year):
1	Bolivia_All.dta	<code>`project'/data/in/</code>	Bolivia – Integrated Household Survey (Encuesta Integrada de Hogares - EIH) – 2015
2	Brasil_All.dta	<code>`project'/data/in/</code>	Brasil – National Household Sample Survey (PNAD) – 2015
3	Ecuador_All.dta	<code>`project'/data/in/</code>	Ecuador – Encuesta de Condiciones de Vida (ECV) – 2013/2014
4	Ethiopia_All.dta	<code>`project'/data/in/</code>	Ethiopia – Socioeconomic Survey (ESS, Wave 3) – 2015/2016
5	Ghana_All.dta	<code>`project'/data/in/</code>	Ghana – Living standards Survey 6 (GLSS) – 2012/2013
6	Uganda_All.dta	<code>`project'/data/in/</code>	Uganda – National Panel Survey (UNPS) – 2015/2016

These data files constitute the main inputs to the cleaning codes (see below) that prepare the final harmonized outputs for the empirical analysis, to be stored in the ``project'/data/out/` folder. In certain cases, some auxiliary inputs were also needed to prepare all variables. These additional components are stored in the ``project'/data/in/Extras` sub-folder and directly referenced by the main data preparation codes.

The table below lists the relevant data files and explains where and how they are used. All such files contain inputs to running the IGROWUP macro developed by the WHO Department of Nutrition and Food Safety for calculating the z-scores and prevalences for nutritional surveys and the direct measurement of nutritional deprivation in Ethiopia and Uganda, where relevant information on child nutritional deprivation

is missing. The full up-to-date contents of the IGROWUP folder is freely accessible through the following GitHub link: https://github.com/unicef-drp/igrowup_update .

Dataset	File name	Country	Function
1	Children_nutri_eth_prev_rc.xls Children_nutri_eth_z_rc.xls Children_nutri_eth_z_rc.dta	ETHIOPIA	The .xls files represent inputs to calculating z-scores and prevalences for nutritional information related to child mortality in Ethiopia using the IGROWUP Stata macro developed by the WHO. The .dta file contains the result of these calculations for measuring nutriton deprivation.
2	Children_nutri_uga_prev_rc.xls Children_nutri_uga_z_rc.xls Children_nutri_uga_z_rc.dta	UGANDA	The .xls files represent inputs to calculating z-scores and prevalences for nutritional information related to child mortality in Uganda using the IGROWUP Stata macro developed by the WHO. The .dta file contains the result of these calculations for measuring nutriton deprivation.

Computational Requirements:

This replication package requires Stata. The analysis was conducted in Stata 18.0.

To run the full package, the replicator should expect the master cleaning do-file to run for about 5 minutes and the master analysis do-file to run for about 2.5 hours.

The execution of the code requires the following Ado package to be installed: "igrowup_restricted.ado". This can be downloaded from: https://github.com/unicef-drp/igrowup_update/tree/master . The relevant contents need to be placed in the Stata/ado/plus folder of the user's computer (type "sysdir" in stata to find the path). (Note that the remaining files from the IGROWUP package are already placed in the project data folder, ``project'/data/in/Extras/igrowup_stata/`. Note additionally, that if using a Mac, please modify all file path entries in igrowup_restricted.ado. Replace \ with / when referencing .dta files.)

Replication procedure

The replication package contains the following elements:

- The main.do file as the single code file to run in order to carry out the replication
- A README.txt file that contains detailed descriptive information and instructions for replication
- “Do” sub-folder with the relevant codes
- “Data” sub-folder with the relevant input / intermediate / output dataset
- “Results” sub-folder with the resulting figures and tables used in the paper

Description of code:

All figures and tables can be generated by running main.do, which is located in the root directory and serves as the single-entry point for the replication. This file sets the project directory, installs all the required packages and runs the cleaning master do-file (“0_dataprep_master.do”) first and the analysis master do-file (“0_analysis_master.do”) second.

Note that the execution of the main.do file requires the replicators to set the main project directory based on the location from where the replication is executed. Currently, the package is placed in a shared Dropbox folder – if the replicator uses that directly, only line 47 needs adjustment. If the replication is executed from another location, line 46 needs to be deleted / commented out, and line 47 appropriately modified.

- 0_dataprep_master.do in the ‘project’/data/in folder: runs 6 different do files
 1. ‘project’/do/cleaning/dataprep_BOL_final.do: compiles the Bolivian data
 2. ‘project’/do/cleaning/dataprep_BRA_final.do: compiles the Brazilian data
 3. ‘project’/do/cleaning/dataprep_ECU_final.do: compiles the Ecuadorian data
 4. ‘project’/do/cleaning/dataprep_ETH_final.do: compiles the Ethiopian data
 5. ‘project’/do/cleaning/dataprep_GHA_final.do: compiles the Ghanaian data
 6. ‘project’/do/cleaning/dataprep_UGA_final.do: compiles the Ugandan data
- 0_analysis_master.do in the ‘project’/data/in folder: executes the analysis and runs 2 do files:
 1. 1a_simulations_dataset.do : runs different hypothetical scenarios for poverty analysis
 2. 1a_simulations_analysis.do : analysis the results of simulated poverty data

All country-specific do-files used to prepare the master data contain two separate parts. The first part contains the code that were used to merge the original source inputs to create the respective COUNTRYNAME_All.dta files contained in the “Data” sub-folder. (This part of the code is currently commented out as it is not needed for the replication.) The second part of the code contains the detailed variable transformations that were used to arrive at the cross-country harmonized and customized subset of variables that are use int common analysis. (This part is required for the replication.)

Instructions for replicators:

- Download the data and request access if needed.
- The project directory is automatically set based on Stata’s working directory. Before running main.do, make sure that Stata’s working directory is the root folder of the replication package.

If this is not the case, edit lines 46-47 `[global project "`r(db)'/ADD_NAME/"]` by replacing only the string after `global project` with the full path to your project directory (e.g., `global project "/Users/.../ReplicationPackage"`).

- Run `main.do` to generate all figures and results.

Code Locations for Figures and Tables

Most figures are produced by the file 0_analysis_master.do. They are clearly indicated, appear at the following lines, and stored in the listed output files (tables in the results>tab folder, and figures in the results>fig folder):

Figure/Table #	Line Number	Output file
Figure 1 Panel A	42	MNM_scatter_190_33.pdf
Figure 1 Panel B	79	MNM_scatter_cstar_190_33.pdf
Figure 2	137	coeffs_cstar_190_33.pdf
Figure 3 Panel a	429	fig_correlation_190_33_all.pdf
Figure 3 Panel b	442	fig_rerankings_190_33_all.pdf
Figure 3 Panel c	455	fig_budgetr_190_33_all.pdf
Figure 4 Panel a	547	H_cneg100_190_33.pdf
Figure 4 Panel b	547	H_cneg50_190_33.pdf
Figure 4 Panel c	547	H_c0_190_33.pdf
Figure 4 Panel d	547	H_c50_190_33.pdf
Figure 4 Panel e	547	H_c100_190_33.pdf
Figure 4 Panel f	547	H_cM_190_33.pdf
Figure 4 Panel g	547	H_cNM_190_33.pdf
Figure 4 Panel h	547	H_cMNM_190_33.pdf
Figure 5 Panel a	976	depthx_corr1.pdf
Figure 5 Panel b	989	depthx_rr1.pdf
Figure 5 Panel c	1002	depthx_budget1.pdf
Figure A1 Panel A	429	fig_correlation_190_33_exc_outliers.pdf
Figure A1 Panel B	442	fig_rerankings_190_33_exc_outliers.pdf
Figure A1 Panel C	455	fig_budgetr_190_33_exc_outliers.pdf
Figure A2	367	fig_difflevels_190_33_all.pdf
Figure A3 Panel A	429	fig_correlation_320_20_all.pdf
Figure A3 Panel B	442	fig_rerankings_320_20_all.pdf
Figure A3 Panel C	455	fig_budgetr_320_20_all.pdf

The following figures are produced by the 1b_simulations_analysis.do file called by the 0_analysis_master.do file. They are listed here below:

Figure/Table #	Line Number	Output file
Figure 6 Panel a	174	fig_budget_sim_scenario_2.pdf
Figure 6 Panel b	174	fig_budget_sim_scenario_3.pdf
Figure 6 Panel c	174	fig_budget_sim_scenario_4.pdf
Figure 6 Panel d	174	fig_budget_sim_scenario_6.pdf
Figure 6 Panel e	174	fig_budget_sim_scenario_10.pdf
Figure 6 Panel f	174	fig_budget_sim_scenario_15.pdf
Figure 6 Panel g	174	fig_budget_sim_scenario_16.pdf
Figure A4 Panel a	194	fig_corr_sim_scenario_2.pdf
Figure A4 Panel b	194	fig_corr_sim_scenario_3.pdf
Figure A4 Panel c	194	fig_corr_sim_scenario_4.pdf
Figure A4 Panel d	194	fig_corr_sim_scenario_6.pdf
Figure A4 Panel e	194	fig_corr_sim_scenario_10.pdf
Figure A4 Panel f	194	fig_corr_sim_scenario_15.pdf
Figure A4 Panel g	194	fig_corr_sim_scenario_16.pdf
Figure A5 Panel a	214	fig_reranks_sim_scenario_2.pdf
Figure A5 Panel b	214	fig_reranks_sim_scenario_3.pdf
Figure A5 Panel c	214	fig_reranks_sim_scenario_4.pdf
Figure A5 Panel d	214	fig_reranks_sim_scenario_6.pdf
Figure A5 Panel e	214	fig_reranks_sim_scenario_10.pdf
Figure A5 Panel f	214	fig_reranks_sim_scenario_15.pdf
Figure A5 Panel g	214	fig_reranks_sim_scenario_16.pdf