Readme for replicating *The Welfare Cost of Drought in Sub-Saharan Africa* by Gascoigne et al. (2024)

For question on the package please contact: Ruth Hill (rhill@worldbank.org) or Katja Vinha (kvinha@worldbank.org)

Replication

Introduction

The reproducibility package described here replicates the analysis in *The Welfare Cost of Drought in Sub-Saharan Africa* by Gascoigne et al. (2024). The document can be found at: Gascoigne et al - 2024 - The Welfare Cost of Drought in Sub-Saharan Africa.

Citation of paper

Gascoigne,Jon; Baquie,Sandra; Vinha,Katja Pauliina; Skoufias,Emmanuel; Calcutt,Evie Isabel Neall; Kshirsagar,Varun Sridhar; Meenan,Conor; Hill,Ruth. *The Welfare Cost of Drought in Sub-Saharan Africa (English).* Policy Research working paper; no. WPS 10683; Washington, D.C.: World Bank Group. http://documents.worldbank.org/curated/en/099325301292478621/IDU1ae0eac0e145d214c6218002156b672eb8155

Computational specifications

The replication requires Excel and Stata.

The code was run on Stata 16.1 on a machine with 16 GB of RAM and an Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz processor. It takes about 40 hours to complete the main do file.

Detailed Instructions

In order to replicate the tables and figures in the paper:

1. Copy the folder "**Drought_Welfare_SSA**" with all the files and folders intact to where you want to run it from

2. Open the folder "**Drought_Welfare_SSA**" and open *Drought_paper_analyses.do*

3. Change line 14 to indicate where **Drought_Welfare_SSA** folder is located.
   Line 14: local main_directory_name **C:/to path**

4. Execute the code which will (automatically) do the following:
   a. Install the following packages if not already installed (lines 20 to 28)
      i. geonear
      ii. fileList
      iii. carryforward
      iv. labutil
      v. distinct
      vi. estout

vii. tabout
viii. outreg2
ix. grstyle

b. The code will execute the following 6 steps. No additional input from the replicator required.

Step 1: Prepare the household surveys for analysis. The inputs for these are in *Input_data->Household_survey->`country'->Dataprep,* where 'country' refers to the one of the nine countries in the study (Ethiopia, Lesotho, Malawi, Mauritania, Mozambique, Niger, Nigeria, Zambia, Zimbabwe). There are a differing number of datasets with varying names depending on the country in question and how preprocessed the data were when obtained by the team.

Step 2: Merge the household surveys (from Step 1) to weather hazards and create country-specific datasets used for pooling. This step is data- and time-intensive.

Step 3: Create the harmonized, pooled datasets used for analyses based on Step 2.

Step 4: Create a historical hazard dataset and it requires that both Steps 2 and 3 have been executed prior.

Step 5: Run the regressions and analyses including the historical simulations based on steps 2, 3 and 4.

Step 6: Create the simulations for Malawi from the AIR simulated weather dataset, based on steps 2, 3, 4 and 5. This step is also very time intensive, taking about 20 hours to run.

Replicated tables and figures

The input for the tables and figures used in the paper are in *Results-> Paper* and are labelled by their figure/table number in the paper. Most of the outputs are Excel tables but some are text files. While all the relevant information is in each file, some require additional formatting to correspond to those in the paper.

For table 3.2 and figures 4.1 the livelihood zones are as follows LZ1 refers to maize, LZ2 refers to agropastorals, LZ3 refers to pastorals, LZ4 refers to roots and lowland, and LZ5 refers to highland. In figure 4.1 chirps refers to rainfall, and swi to soil moisture. Similarly in all tables headings LZ:1 refers to maize livelihood zone, LZ:4 to roots and lowland, and LZ:5 to highland.

The columns in table46 need to be reorganized to match those in the paper.

In tables 4.7 and 4.8 the total consumption columns come from table 4.1 and 4.2, respectively and are not included in table47_r20 or table48_r20.

Excel files table51_r20 and table 52_r20 contain separate tabs for livelihood zones using growing season based measures (maize, highland, roots and lowland), for pastorals, and for agropastorals.

Data Sources and Availability

There are four different types of data used in the analysis:

(A) exposure data;
(B) climate/weather data;
(C) household consumption surveys;
(D) simulated weather data.

Data in A and B are based on for most part on publicly available data (Tables A and B). The publicly available data was downloaded and the datasets were processed in RStudio software to create gridded layers, defined by a 0.05° by 0.05° 'Uniform Resolution Grid (URG)', and the hazard data on a consistent decadal (10-day) or monthly resolution depending on the minimum available temporal resolution of the source data. Categorical data including administrative zone, livelihood zone, agroecological zone and WRSI were extracted at the centroid of the uniform resolution grid, continuous data including population, crop coverage, travel time to markets, temperature, precipitation, evapotranspiration, soil moisture, and vegetation condition were linearly interpolated to the URG centroid. Hazard data were aggregated to 10-day decadal intervals where the source data was provided at finer resolutions and mapped to the nearest decadal interval where source data were only available at coarser temporal resolution. These data were compiled by the Center for Disaster Protection from the specified sources. The scripts for these compilations are not available and not part of the reproducibility package. These gridded data are available at the World Bank's Data for Development Hub.

Data in C are household consumption surveys provided by country poverty economists in the World Bank's Poverty and Equity Global Practice. Table C lists names of the household surveys used. The versions with the consumption aggregate **are not publicly available** and not included in the package. Ruth Hill (rhill@worldbank.org) can be contacted for details.

Data in D are from a set of stochastic catalogs of synthetic simulations for Malawi at a spatial resolution of 0.05° produced by Verisk's AIR Worldwide (AIR) for this work and **are not publicly available.** In total 10,000 simulations are used. These are housed in the World Bank's Data for Development Hub (https://datacatalog.worldbank.org/int/search/dataset/0066871/simulated_dekadal_soil_water_index_swi) but are not publicly accessible and cannot be requested.

Data statement

- All data are publicly available.
- Some data cannot be made publicly available.
- No data can be made publicly available

Statement about Rights

I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Table A Information on exposure data

Data set	Source of data	Folder location
		NOTE: Country abbreviations used for country specific datasets: ETH: Ethiopia, LSO: Lesotho, MWI: Malawi, MRT: Mauritania, MOZ: Mozambique, NER: Niger, NGA: Nigeria, ZMB: Zambia, ZWE: Zimbabwe
GADM – Database of Global Administrative Areas (Admin0–Admin6)	https://datacatalog.worldbank.org/search/dataset/0066860/gridded_global_administrative_areas_gadm_for_africa	./Input_data>Remote_Sensed>SSA>gadm_admin_urg005_mapping>`country_abbreviation'_gadm_admin_urg005_mapping
GAUL – Global Administrative Unit Layers (Gaul0–Gaul1)	https://datacatalog.worldbank.org/search/dataset/0066859/Gridded-Global-Administrative-Unit-Layers--GAUL--for-Africa	./Input_data>Remote_Sensed>SSA>country_mapping>`country_abbreviation'_gaul1_asap_urg005_mapping
Crop coverage – Copernicus Global Land Cover Layers	https://datacatalog.worldbank.org/search/dataset/0066860/gridded_global_administrative_areas_gadm_for_africa	./Input_data>Remote_Sensed>SSA>crop_coverage>country_mapping>'country_abbreviation'_CROPCOV19_mapping
Crop calendar	https://datacatalog.worldbank.org/int/search/dataset/0066842/crop_calendar_for_africa	./Input_data>Remote_Sensed>SSA>merge_variable_data>crop_calendar_gaul1 (subnational) ./Input_data>Data_Integration>SSA>crop_calendar_gaul_asap1
Travel Time to Markets – International Food Policy Research Institute (IFPRI)	https://datacatalog.worldbank.org/int/search/dataset/0066865/gridded_travel_time_to_market_for_africa	./Input_data>Remote_Sensed>SSA>time_to_market>country_mapping>`country_abbreviation '_ttm_urg005_mapping
Global Agro-Ecological Zones (GAEZ)	https://datacatalog.worldbank.org/int/search/dataset/0066857/gridded_global_agroecological_zones_gaez_for_africa	./Input_data>Remote_Sensed>SSA>agro_ecological_zone>country_mapping>`country_abbreviation'_aez_urg005_mapping
Livelihood zones	https://datacatalog.worldbank.org/int/search/dataset/0066861/gridded_livelihood_zones_for_africa	./Input_data>Remote_Sensed>SSA>livelihood_zone>country_mapping>`country_abbreviation'_livelihood_urg005_db

Source: World Bank compilation.

Table B Information on hazard data

Data set	Source of data	Folder location
		NOTE: Country abbreviations used for country specific datasets: ETH: Ethiopia, LSO: Lesotho, MWI: Malawi, MRT: Mauritania, MOZ: Mozambique, NER: Niger, NGA: Nigeria, ZMB: Zambia, ZWE: Zimbabwe
Precipitation	https://datacatalog.worldbank.org/int/search/dataset/0066858/gridded_dekadal_precipitation_data_chirps_for_africa	./Input_data>Remote_Sensed>SSA>CHIRPS>CHIRPS_country_dekadal_xy_output
Evapotranspiration	https://datacatalog.worldbank.org/int/search/dataset/0066863/gridded_3month_standardized_precipitationevapotranspiration_index_spei_for_africa	./Input_data>Remote_Sensed>SSA>SPEI3
Water availability for crops	https://datacatalog.worldbank.org/int/search/dataset/0066866/gridded_dekadal_water_requirement_satisfaction_index_wrsi_for_africa	./Input_data>Remote_Sensed>SSA>WRSI
Soil water index	https://datacatalog.worldbank.org/int/search/dataset/0066864/gridded_dekadal_soil_water_index_swi_for_africa	./Input_data>Remote_Sensed>SSA>SWI
NDVI	https://datacatalog.worldbank.org/int/search/dataset/0066862/gridded_monthly_normalized_difference_vegetation_index_ndvi_for_africa	./Input_data>Remote_Sensed>SSA>NDVI>MODIS13C2_output_xyz

Source: World Bank compilation

Table C Consumption and expenditure surveys

Country	Year	Survey
Ethiopia	2010/2011	Household Income Consumption and Expenditure Survey
	2015/2016	Household Income Consumption and Expenditure Survey
Lesotho	2017/2018	Continuous Multipurpose Household Survey/ Household Budget Survey
Malawi	2010/2011	Third Integrated Household Survey
	2016/2017	Fourth Integrated Household Survey
	2019/2020	Fifth Integrated Household Survey
Mauritania	2014	Enquête Permanente des Conditions de Vie
	2019/2020	Enquête Permanente des Conditions de Vie
Mozambique	2014	Inquérito aos Agregados Familiares sobre Orçamento Familiar (IOF)
Niger	2018	Enquête Harmonisée sur le Conditions de Vie des Ménages
Nigeria	2011	General Household Survey
	2013	General Household Survey
	2016	General Household Survey
Zambia	2015	Living Conditions Monitoring Survey
Zimbabwe	2017	Poverty, Income, Consumption and Expenditure Survey

Folder structure

The reproducibility package is as follows. The structure must be kept intact for the main script to execute properly. Within the main folder Drought_Welfare_SSA there are 4 different folders.

1. **Do_Files**
   Within the Do_Files folder there is a country specific folder for each country in the folder Country. Here are the do files used to prepare the data *(`Country'_dataprep.do*), to integrate the household survey with the hazard data (*HH_Survey_Hazard_Integration_`Country'.do*) as well as to produce the country specific simulated poverty rate and poverty gap figures *(`Country'_EP_curves_prg.do*). There is also a country folder named Pooled which contains the do files for working with the pooled dataset including all the code for harmonizing the data (*Harmonized_data.do*) and performing the analyses (*Regressions_pooled.do*). The other scripts in the folder are ones called by Regressions_pooled.do to perform certain actions.

   The folder General contains the scripts used for extracting the hazard and contextual data for the households in the household surveys. These are located in *General ->Programs->Data_Integration* and they get called on by *HH_Survey_Hazard_Integration_`Country'.do*.

2. **Input_Data**
   Input_Data has four different folders. These are raw or intermediate data used to feed into the *HH_Survey_Hazard_Integration_`Country'.do* script as well as the country level output datasets from *HH_Survey_Hazard_Integration_`Country'.do* which are used in the analysis stage. Input data needs to be considered as both an input and output data folder.

   1. Household_Survey

      Household_Survey has the household consumption / expenditure datasets which are used in *`Country'_dataprep.do* to create the household dataset used as an input to *HH_Survey_Hazard_Integration_`Country'.do*. The source for the underlying data used in *`Country'_dataprep.do* are given in Table C and the versions used are in the World Bank Microdata repository. What will also be housed here after the code has been executed are the final datasets produced by *HH_Survey_Hazard_Integration_`Country'.do* which are used as inputs in the analysis scripts.

      The folder has a **subfolder for each country**. The raw household survey used for the analysis will be here as well as the integrated datasets once they are generated. Within each country folder, the *Dataprep* folder contains all the required datasets needed to extract the variable from the household surveys necessary for the analysis.

      In the **Pooled folder** the harmonized integrated country datasets for all radii considered (10, 20, 50 km) will be placed. For those countries where administrative boundaries are used there will be three equal datasets. Also, the combined dataset pooled across all 9 countries will be stored here. These data are the inputs into the analysis script.

2. Remote_Sensed
   In the subfolder SSA there are 11 folders, one of each of the hazards or contextual variables in their gridded form and folder *merge_variable_data* with the start and end dates for growing season and rainy season periods to be used in *HH_Survey_Hazard_Integration_`Country'.do*. See information in Tables A and B for the source of the data and the specific folder path for these data.

Folder	Gridded contents
agro_ecological_zone	Agro-ecological zones
CHIRPS	Rainfall
country_mapping	Administrative boundaries (GAUL)
crop_coverage	Land coverage in crops
Gadm_admin_urg005_mapping	Administrative boundaries (GADM)
Livelihood_zone	Livelihood
NDVI	NDVI
SPEI3	SPEI
SWI	Soil moisture
Time_to_market	Time to population agglomerations
WRSI	WRSI

3. Data_Integration
   In the Country subfolder there is a folder for each country which includes the household survey with GPS coordinates or with administrative area identifiers used by *HH_Survey_Hazard_Integration_`Country'.do*. These are produced in the data preparation stage for each country. If the matching is done by administrative areas (Ethiopia and Zimbabwe) an additional csv file is located here for the proper spatial matching of the households and hazard/contextual variables.

   In the SSA folder are the crop calendar data for each administrative unit (*crop_calendar_gaul_asap1*), main crop for each administrative unit (*FAOSTAT_data_area_GAUL*), and dataset for expanding datasets to cover all months from January 2000 to December 2020 (*year_dekad_2000_2020*).

4. Simulated
   For each country a separate folder with stochastic catalogs of synthetic simulations from AIR (*smoothed.xxx.summary*). We only use data for Malawi. This is also the folder where the datasets generated in the simulations exercise are placed.

3. Processed_Data

There is a separate subfolder for each country. The processed datasets generated by *HH_Survey_Hazard_Integration_`Country'.do* will be located here.

4. Results

The folder contains all the results of the analyses. Within Results, the **folder Paper** has all the relevant figures and tables that are in the paper. The folder for each country has the country-

specific results from the simulation exercises. In the folder Pooled are all the results based on the pooled dataset.

List of tables/figures and programs

Table/Figure	Program *Do_files>Country>Pooled>*	Line	Output file *Results>Paper*	Notes
Table 3.1	N/A			
Table 3.2	historical_hazards.do	447	table32_LZ1_20.txt	For maize
			table32_LZ2_20.txt	For agropastorals
			table32_LZ3_20.txt	For pastorals
			table32_LZ4_20.txt	For roots/lowland
			table32_LZ5_20.txt	For highland
Table 3.3	N/A			
Table 3.4	outcomes_poly_cons_land_s_int.do	75	table34_agropastoral_r20.txt	
	outcomes_poly_cons_land_wt_int.do	88	table34_gs_r20.txt	For maize, roots/lowland, and highland
	outcomes_poly_cons_tlu_wt_int.do	58	table34_pastoral_r20.txt	
Table 3.5	Harmonized_data.do	818	table35_r20.xlsx	
Figure 3.1	N/A			
Table 3.6	outcomes_poly_cons_land_s_int.do	88	table36_agropastoral_r20.txt	
	outcomes_poly_cons_tlu_wt_int.do	72	table36_pastoral_r20.txt	
	outcomes_poly_cons_land_wt_int.do	88	table36_gs_r20.txt	
Table 4.1	outcomes_poly_cons_land_wt_int.do	136,163,185	table41_r20.xls	
Table 4.2	outcomes_poly_cons_land_wt_int.do	142,169,191	table42_r20.xls	
Table 4.3	outcomes_poly_cons_land_wt_int.do	139,165,188	table43_r20.xls	
Figure 4.1 a	outcomes_poly_cons_land_wt_int.do	219	figure41_2_LZ1_anomaly_swi_p_r20.gph	Maize, soil moisture
Figure 4.1 b	outcomes_poly_cons_land_wt_int.do	219	figure41_2_LZ1_anomaly_ndvi_p_r20.gph	Maize, NDVI
Figure 4.1 c	outcomes_poly_cons_land_wt_int.do	219	figure41_2_LZ5_anomaly_swi_p_r20.gph	Highland, soil moisture
Figure 4.1 d	outcomes_poly_cons_land_wt_int.do	219	figure41_2_LZ5_anomaly_ndvi_p_r20.gph	Highland, NDVI
Figure 4.1 e	outcomes_poly_cons_land_wt_int.do	219	figure41_2_LZ4_anomaly_swi_p_r20.gph	Roots/lowland, soil moisture
Figure 4.1 f	outcomes_poly_cons_land_wt_int.do	219	figure41_2_LZ4_anomaly_ndvi_p_r20.gph	Roots/lowland, NDVI
Figure 4.2 a	outcomes_poly_cons_land_wt_int.do	219	figure41_2_LZ1_anomaly_spei3_p_r20.gph	Maize, SPEI
Figure 4.2 b	outcomes_poly_cons_land_wt_int.do	219	figure41_2_LZ1_anomaly_chirps_p_r20.gph	Maize, rainfall

Figure 4.2 c	outcomes_poly_cons_land_wt_int.do	219	figure41_2_LZ5_anomaly_spei3_p_r20.gph	Highland, SPEI
Figure 4.2 d	outcomes_poly_cons_land_wt_int.do	219	figure41_2_LZ5_anomaly_chirps_p_r20.gph	Highland, rainfall
Figure 4.2 e	outcomes_poly_cons_land_wt_int.do	219	figure41_2_LZ4_anomaly_spei3_p_r20.gph	Roots/lowland, SPEI
Figure 4.2 f	outcomes_poly_cons_land_wt_int.do	219	figure41_2_LZ4_anomaly_chirps_p_r20.gph	Roots/lowland, rainfall
Table 4.4	outcomes_poly_cons_tlu_wt_int.do	102,110,116	table44_r20.xls	
Table 4.5	outcomes_poly_cons_land_s_int.do	126,133,139	table45_r20.xls	
Table 4.6	outcomes_poly_cons_land_gps.do	75, 84	table46.xls	Includes also first 3 result columns from table 4.1
Table 4.7	outcomes_poly_disag_cons_land.do	73,99,119	table47_r20.xls	Includes also first 3 result columns from table 4.2
Table 4.8	outcomes_poly_disag_cons_land.do	75, 102, 122	table48_r20.xls	
	residualizing_variation_agro_rs.do	134	table51_r20.xls, agropastoral	
Table 5.1	residualizing_variation_gs_v2.do	119	table51_r20.xls, GS	For maize, roots/lowland and highland zones
	residualizing_variation_rs_v2.do	116	table51_r20.xls, pastoral	
Table 5.2	Harmonized_data.do	845	table52_r20.xlsx, table52_gs	For maize, roots/lowland and highland zones
		865	table52_r20.xlsx, table52_agro	For agropastorals
		882	table52_r20.xlsx, table52_pastoral	For pastorals
	Do_files>Country			
Figure 5.1 a	Ethiopia>Ethiopia_EP_curve_prg.do	233	figure51_base_Ethiopia_r20.jpg	
Figure 5.1 b	Lesotho>Lesotho_EP_curve_prg.do	241	figure51_base_Lesotho_r20.jpg	
Figure 5.1 c	Malawi>Malawi_EP_curve_prg.do	238	figure51_base_Malawi_r20.jpg	
Figure 5.1 d	Mauritania>Mauritania_EP_curve_prg.do	255	figure51_base_Mauritania_r20.jpg	
Figure 5.1 e	Mozambique>Mozambique_EP_curve_prg.do	237	figure51_base_Mozambique_r20.jpg	
Figure 5.1 f	Nigeria>Nigeria_EP_curve_prg.do	240	figure51_base_Nigeria_r20.jpg	
Figure 5.1 g	Zambia>Zambia_EP_curve_prg.do	244	figure51_base_Zambia_r20.jpg	
Figure 5.1 h	Zimbabwe>Zimbabwe_EP_curve_prg.do	226	figure51_base_Zimbabwe_r20.jpg	
Figure 5.2 a	Ethiopia>Ethiopia_EP_curve_prg.do	251	figure52_base_Ethiopia_r20.jpg	
Figure 5.2 b	Lesotho>Lesotho_EP_curve_prg.do	259	figure52_base_Lesotho_r20.jpg	
Figure 5.2 c	Malawi>Malawi_EP_curve_prg.do	256	figure52_base_Malawi_r20.jpg	
Figure 5.2 d	Mauritania>Mauritania_EP_curve_prg.do	273	figure52_base_Mauritania_r20.jpg	

Figure 5.2 e	Mozambique>Mozambique_EP_curve_prg.do	255	figure52_base_Mozambique_r20.jpg
Figure 5.2 f	Nigeria>Nigeria_EP_curve_prg.do	258	figure52_base_Nigeria_r20.jpg
Figure 5.2 g	Zambia>Zambia_EP_curve_prg.do	263	figure52_base_Zambia_r20.jpg
Figure 5.2 h	Zimbabwe>Zimbabwe_EP_curve_prg.do	244	figure52_base_Zimbabwe_r20.jpg
Figure 5.3 a	Malawi>MWI_EP_simulated_rev_data.do	403	figure53a.jpg
Figure 5.3 b	Malawi>MWI_EP_simulated_rev_data.do	419	figure53b.jpg
Appendix			
Table A.1	N/A		
Table B.1	N/A		
Table B.2	N/A		
Table C.1	N/A		
Figure C.1	N/A		
Figure C.2	N/A		
Figure D.1	N/A		
Figure D.2	N/A		
Figure D.3	N/A		