Reproducibility Package README

COLLAPSE AND RECOVERY
How the COVID-19 Pandemic Eroded Human Capital and What to Do about It

Overview

Our primary objective in this reproducibility package is to ensure transparency, accessibility, and replicability of the research findings presented in the book, "COLLAPSE AND RECOVERY: How the COVID-19 Pandemic Eroded Human Capital and What to Do about It." Given the extraordinary importance of the topics explored within the report, it is crucial that we provide a comprehensive package for replicating our analyses and insights.

This reproducibility package serves as a valuable resource for researchers, policymakers, and the wider community who wish to examine, validate, and build upon the findings and methodologies presented in the book. By offering a detailed account of the data, code, and procedures used in our research, we aim to foster greater understanding and engagement with the critical issues surrounding human capital, crises, and policy responses.

Moreover, as the pandemic's consequences continue to unfold and impact societies worldwide, it becomes imperative to have a clear, replicable framework for evaluating and addressing these challenges. This reproducibility package not only facilitates the reproducibility of our work but also empowers others to contribute to the ongoing discourse on human capital and crisis management. Ultimately, the dissemination and accessibility of this package are essential steps toward collectively charting a more resilient and equitable future in the face of global challenges.

Getting Started

Downloading the reproducibility package
The first step is to download the reproducibility package. Users can choose to run a compact version that starts from intermediate data and has a size of 75 MB (135 MB uncompressed) or the complete reproducibility package that has a compressed size of 21 GB (and around 320 GB uncompressed). Replicators running the complete reproducibility package should note that it generates intermediate files for a total of 700 GB, so they should attempt to run the complete package in computers that have at least this amount of disk space.

The reproducibility package files are available in these links:

- [Reproducibility package starting from intermediate data](#)
- [Raw data](#)

The complete reproducibility package requires both files, while the compact version requires only the first.

Running the code
Replicators who run the compact reproducibility package should unzip the compressed file (reproducibility package starting from intermediate data), open the main do-file (*0. Master dofile*), set the user root path in line 25, and run it.

Replicators who run the complete reproducibility package should follow these instructions after downloading both reproducibility package files:

1. Unzip the folder with the intermediate data reproducibility package
2. Unzip the folder *DataWork* from the compressed raw data in the folder you extracted in (1), **replacing** the existing *DataWork* folder
3. In the main do-file (*0. Master dofile*), activate lines 95-150. These lines run sub do-files executing the data cleaning and construction that is skipped in the compact reproducibility package.
4. Set the user root path in line 25 of the main do-file and run it.

Installation

To reproduce the results in this replicability package, you will need to use Stata version 17.0 MP—Parallel Edition. However, it is worth noting that the code has also been successfully executed in Stata 15.0 MP—Parallel Edition. If you do not already have Stata 16.0 or a more recent edition installed, you can obtain it from the official Stata website: StataCorp.

Usage

You have the flexibility to execute the entire do file or select specific sections. However, it is essential to run the do file sequentially. If you choose to run only one section, ensure that the preceding sections have already been executed. This is a brief description of the main do-file's content:

PART 0: USER INPUTS
This section sets the global main folder path for the reproducibility package.

PART 1: PREPARE FOLDER PATHS
This section defines global variables for various folder paths related to data and chapters. It also creates subfolder paths for individual countries and chapter outputs.

PART 2: FIGURES SET UP
This part contains commands to set up the graph settings and define global variables for colors and schemes used in figures.

PART 3: RUN SELECTED CODE
This is the main section where code execution takes place. It runs specific do-files for data cleaning, enrollment indicators, labor indicators, merging datasets, and reproducing figures. The code is organized based on different chapters (Chapter 2, Chapter 3, and Chapter 4), and specific do-files for generating figures and text numbers are executed for each chapter.

Reproducibility Steps

For this report, we cleaned and harmonized micro-level data from 16 different countries. This processed data serves as the core input for each chapter. Within the respective chapter folders, you will find the cleaned data along with the corresponding do files required to replicate the figures and calculations presented in the report. The 'Datawork' folder, you will find the microdata processing for each country. The country-specific codes follow the following structure:

1. ***Cleaning:***
   This code is specific to each country, and it processes the raw data to create a harmonized database with variables related to employment and school enrollment.

2. ***Labor indicators:***

This code is designed to standardize employment indicators for each country. It involves several key steps:

❖ **Creation of Standardized Outcomes:**
This section involves creating various indicators related to employment categories, such as employed (emp), unemployed (unemp), inactive (inac), enrolled (enroll), and more. These indicators are generated based on specific criteria within the dataset.

❖ **Creation of Totals:**
The code calculates and saves totals for various employment-related variables, both overall and disaggregated by gender and age group.

❖ **Creation of Rates:**
This section calculates employment rates, unemployment rates, inactivity rates, NEET (Not in Employment, Education, or Training) rates, and enrollment rates based on the totals saved in the dataset.

❖ **Clean-Up:**
This section involves labeling variables for clarity and saving the harmonized dataset.

3. *Enrollment indicators:*
This code standardizes enrollment indicators based on the cleaned dataset. The code creates standardized outcomes, such as binary enrollment indicators, and standardized covariates, including gender, age groups, and education levels. It also calculates enrollment rates for various categories, such as age, gender, and education levels. Afterward, the code merges datasets to consolidate enrollment rates across different categories. It adds a country variable, labels variables for clarity, and finally saves the harmonized dataset for further analysis.

In the 'Datawork' folder, you will also find two additional folders that do not correspond to specific country analyses: 'Harmonized' and 'Other.' After executing the code for individual countries as described earlier, two harmonized data files are generated for each country and stored within the 'Data' subfolder of the 'Harmonized' directory.

The 'Harmonized' folder houses these harmonized enrollment and labor data, which are essential for implementing the primary methodology used to assess the pandemic's impact while accounting for trends and seasonal patterns. To facilitate this analysis, we have created a program named '0_PredictedFunction.do.'

This program is subsequently utilized in the ' 2_Apply_Methodology_Observed_vs_Predicted.do' do file to generate predicted values for employment and enrollment rates. Consequently, it is crucial to execute these do files sequentially and in tandem. This ensures that Stata properly stores the function. Failure to do so may result in an error when running the predicted function. A more detailed description of the methodology can be found in the Annex 4.1. of the report.

Data

In the following table, you will find information about both country sources and aggregate data sources used in this reproducibility package. For country sources, the table specifies the name of the survey used, its frequency, the periods utilized, and whether the source is publicly accessible or confidential.

	Country/Dataset	Survey Name	Region	Frequency	Periods	Type
Country sources	Argentina	Encuesta Permanente de Hogares (EPH)	LAC	Quarterly	2012Q1 - 2021Q4	Publicly available
	Brazil	Continuous PNAD - Continuous National Household Sample Survey	LAC	Quarterly	2012Q1 - 2021Q4	Publicly available
	Bulgaria	Eurostat	ECA	Quarterly	2012Q1 - 2021Q4	Publicly available
	Colombia	Gran Encuesta Integrada de Hogares - GEIH	LAC	Monthly	2012M1 – 2022M1	Publicly available
	Ethiopia	Urban Employment and Unemployment Survey (UEUS)	SSA	Yearly	2012, 2014, 2015, 2016, 2018, 2020 (pre-pandemic)	Confidential
		Labor Force and Migration survey (LFMS)		Yearly	2013, 2021	Confidential
	India	Consumer Pyramids Household Survey CMIE	SA	Every 4 months	2017T1-2021T3	Confidential
	Indonesia	National Socioeconomic Survey (SUSENAS)	EAP	Yearly	2015-2021	Confidential
	Jordan	Tables of Employment and Unemployment - Department of Statistics, Amman (Jordan).	MENA	Quarterly	2012Q1 - 2021Q4	Publicly available
	Mexico	Encuesta Nacional de Ocupación y Empleo (ENOE)	LAC	Quarterly	2012Q1 - 2021Q4	Publicly available
	Morocco	ACTIVITÉ, EMPLOI ET CHÔMAGE (TRIMESTRIEL)	MENA	Quarterly	2012Q1 - 2021Q5	Reports Publicly available
	Pakistan	Labor Force Survey (LFS)	SA	Quarterly	6 waves (2012-13, 2013-14, 2014-15, 2017-18, 2018-19 and 2020-21)	Confidential
	Phillippines	Labor Force Survey (LFS)	EAP	Quarterly	2012Q1 - 2021Q2	Confidential
	South Africa	Quarterly Labour Force Survey	SSA	Quarterly	2012Q3 - 2021Q3	Publicly available
	Turkey	Labor Force Survey (LFS)	ECA	Yearly	2012-2020	Confidential Aggregate tables:Publicly available
	Uruguay	Encuesta Continua de Hogares	LAC	Monthly	2012Q1 - 2021Q2	Publicly available
	Vietnam	Labor Force Survey (LFS)	EAP	Quarterly	2012Q1-2021Q4	Confidential
Aggregate data sources	Immunization - UNICEF	WUENIC	Global			Publicly available
	UIS Survey of National Education	Survey of National Education Responses to COVID-19 School Closures		3 rounds		Publicly available
	UIS Global monitoring of school closures	Global Monitoring of School Closures Caused by the COVID-19 Pandemic - Database of duration of school closures				Publicly available
	World Bank's Databank	World Development Indicators				Publicly available
	WB Worldwide Governance Indicators	Worldbwide Governance Indicators				Publicly available

Here you can find instructions on how to download the non confidential data in the report.

- ❖ **Argentina**: Encuesta Permanente de Hogares (EPH)
- ❖ **Brazil**: Continuous PNAD - Continuous National Household Sample Survey

- ❖ **Bulgaria**: Eurostat
  - o [Employment by sex, age and educational attainment level (1000):](#)
  - o [Population by sex, age and educational attainment level (1 000)](#)
  - o [Young people neither in employment nor in education and training (NEET), by sex and age - quarterly data](#)
- ❖ **Colombia**: [Gran Encuesta Integrada de Hogares - GEIH](#)
- ❖ **India**: [Consumer Pyramids Household Survey CMIE](#) (Available for purchase)
- ❖ **Jordan**: [Department of Statistics, Amman (Jordan)](#)
  - o Table 2.1: Jordanian Population Age 15+ Years by Educational Level, Sex and Broad Age Groups (Percentage Distribution).
  - o Table 3.1: Jordanian Population Age 15-24 Years by Activity Status and Educational Enrollment status (Percentage Distribution).
  - o Table 5.2: Jordanian Employed Persons Age 15+ Years by Sex, Major Age Group and Educational Level (Percentage Distribution).
- ❖ **Mexico:** [Encuesta Nacional de Ocupación y Empleo (ENOE)](#)
- ❖ **Morocco:** [ACTIVITÉ, EMPLOI ET CHÔMAGE (TRIMESTRIEL)](#) (reports, data extracted manually)
- ❖ **South Africa:**
  - o [Quarterly Labour Force Survey](#) and
  - o [Labour Market Dynamics in South Africa](#) (for wages)
- ❖ **Türkiye**: [Labour Force Statistics (2014 and after)(M)](#) database. The following measurements were used:
  - o Labour Force (Thousand),
  - o Employment (Thousand),
  - o Not In Labour Force Thousand.
- ❖ **Uruguay:** [Encuesta Continua de Hogares](#)
- ❖ **Other:** Auxiliary aggregate sources for cross country indicators
  - o [WUENIC](#)
  - o [Survey of National Education Responses to COVID-19 School Closures](#)
  - o Global Monitoring of School Closures Caused by the COVID-19 Pandemic - [Database of duration of school closures](#)
  - o [World Development Indicators](#)
  - o [Worldbwide Governance Indicators](#)

The raw data file of the reproducibility package contains all the publicly available data. For countries where confidential data was used, we provide the cleaning do files (Read only) and the resulting harmonized data. The raw data for the Publicly available data sources described above is approximately 294 GB.

Results

Please be aware that, owing to updates performed by IBGE in Brazil and INEGI in Mexico, certain results presented in the replicated materials exhibit minor discrepancies when compared to the values originally reported in the source document. The initial results in the report were derived using PNAD and ENOE data acquired in 2022. To provide an updated version of the graphs, this package utilizes the most recent data available for all countries, which was obtained in 2023. Specifically, in the cases of Mexico [ENOE](#) and Brazil [PNAD Continua](#), variations have occurred due to reweighting procedures implemented

by the respective statistical institutes. It is important to emphasize that all other estimations included in this package remain consistent with those presented in the original report. The overarching trends and relative magnitudes of the effects remain unchanged, and the core messages remain valid.

In this section, you will find a comprehensive list of the figures that have been successfully reproduced. Figures that don't reproduce exactly due to changes in the data posterior to the publication are marked with an asterisk. Please note that figures not listed here, but appearing in the report, were generated from secondary sources and other published studies. These figures are not part of the primary data analysis conducted for this report and are therefore not included in this reproducibility package. The original sources for these figures are appropriately cited within the report text for reference.

- o **Figure 2.3** By 2021, coverage of essential childhood vaccines had yet to recover fully in many regions from its decline during lockdowns.
- o **Figure 2.4** Pre-primary attendance has not recovered from the pandemic in many countries.
- o Figure 3.1 Figure 3.1 Globally, an average school-age child lost about one year of in-person schooling.

- o **Figure 3.2** COVID-19 school closures had limited impacts on dropouts in middle-income countries but negative impacts in lower-income countries.
- o **Figure 3.3*** Dropout rates are higher for households with low education levels.
- o **Figure 3.4** For 30 days of school closures, students lost 34 days of learning.
- o **Figure 3.5** Learning losses were higher in countries with lower GDP per capita after controlling for length of school closures.
- o **Figure 3.6** Regions vary in the Learning-Adjusted Years of Schooling (LAYS) they lost due to the pandemic.
- o **Figure 3.7** Countries that had similar Learning-Adjusted Years of Schooling (LAYS) before the pandemic had vastly different experiences with learning losses.
- o **Figure 3.8** Figure 3.8 Countries not losing many total Learning-Adjusted Years of Schooling (LAYS) may have lost much in terms of pre-pandemic shares of LAYS (and vice versa)
- o **Figure B3.2.1** There is no systematic relationship between the length of school closures and log GDP per capita and an indicator of governance effectiveness.

- o **Figure 4.1** Worldwide, employment fell sharply during the pandemic.
- o **Figure 4.2*** Youth employment declined sharply during the pandemic.
- o **Figure 4.3*** In many countries, the employment losses of youth during the pandemic were compounded by declines in wages.
- o **Figure 4.4*** School enrollment increased in some countries and declined in others during the pandemic.
- o **Figure 4.5*** The share of youth who were Not in Education, Employment, or Training (NEETs) increased sharply in some countries during the pandemic.
- o **Figure 4.6** In some countries, both youth and adult employment have not recovered from the pandemic.
- o **Figure 4.7** In some countries, adult employment has recovered from the pandemic, but youth employment has not or has recovered less.
- o **Figure 4.8*** In some countries, both youth and adult employment have recovered from the pandemic.

- **Figure 4A.1** The importance of removing trends and seasonal effects for the estimation of the effects of the pandemic on employment and other outcomes.
- **Figure B4.1.1** Three scenarios show how short-term employment losses can affect a young person's future wages.

This reproducibility package shares the underlying data and associated code required to replicate all figures in the report. Please note that figures 3.6, 3.7, and 3.8 build on country-level data from the Human Capital Index that have not yet been publicly released. Consequently, this package shares a country-level dataset which anonymizes country names (while retaining information on World Bank regions and income groups). As a result, the code to label country names will not work for figures 3.7 and 3.8 as produced in the report. Once the full country-level data are ready for public release, they will also be updated as part of this reproducibility package.

Contributing

Contributions to our project are highly encouraged and welcome. If you're interested in improving or enhancing this reproducibility package, you can contribute by identifying and fixing bugs, proposing new features, or suggesting improvements. Your contributions will play a crucial role in enhancing the quality and utility of this reproducibility package, and we appreciate your support in making this project even better.

License

This reproducibility package utilizes Stata Version 17.0 MP—Parallel Edition for its execution. Please note that the primary licensing and copyright terms for the Stata software apply to its usage within this package. For more details regarding Stata's licensing and copyright information, we recommend referring to Stata Corporation's official documentation and terms of use.

Acknowledgments

Contact

For any inquiries or assistance related to this reproducibility package, please feel free to reach out to Margarita Isaacs at misaacsprieto@worldbank.org and Diego Parra at dparraalvarez@worldbank.org. We are here to help and address any questions you may have regarding the contents and utilization of this package. Your feedback and inquiries are greatly valued.