README

Data and Code for "Measuring Green Jobs: A New Database for Latin America and Other Regions"

by Kelly Y. Montoya, Emmanuel Vazquez, and Hernan Winkler[1]
(May 2024)

Overview

This document provides information needed to replicate the results in "Measuring Green Jobs: A New Database for Latin America and other Regions".

The folders included in this replication package are the following:

- **Data**
- **Do_files**
- **Outputs**

Data, Do_files, and Outputs should be in the same folder (e.g. the folder they were archived in) to run the project do-files below.

Computational requirements

Software Requirements: Stata version 14 or higher is required. The guidelines below, sourced from https://www.stata.com/products/compatible-operatingsystems/, outline the following computational requirements:

Platforms

Stata for Windows®

- Windows 11*
- Windows 10 *
- Windows Server 2022, 2019, 2016, 2012R2 *

* Stata requires 64-bit Windows for x86-64 processors made by Intel® or AMD (Core i3 equivalent or better)
Find out if your OS is 64-bit compliant.

Stata for Mac®

- Mac with Apple Silicon or Intel processors
- macOS 11.0 (Big Sur) or newer for Macs with Apple Silicon and macOS 10.13 (High Sierra) or newer for Macs with Intel processors

Stata for Linux

- Any 64-bit (Core i3 equivalent or better) running Linux
- Minimum requirements include the GNU C library (glibc) 2.17 or better and libcurl4
  - Check the output of **ldd -v** within a terminal
- For xstata, you need to have GTK 2.24 installed

Hardware requirements

Package	Memory	Disk space
Stata/MP	4 GB	2 GB
Stata/SE	2 GB	2 GB
Stata/BE	1 GB	2 GB

Stata for Linux requires a video card that can display thousands of colors or more (16-bit or 24-bit color)

[1] Montoya and Winkler are with the World Bank Poverty and Equity Global Practice, and Vazquez is with CEDLAS (IIE, FCE) - Universidad Nacional de La Plata, Argentina. Correspondence author: Hernan Winkler (hwinkler@worldbank.org).

Packages needed for tables and figures: outreg2. They are included in the ado folder (Do_files/ado).

Memory and Runtime Requirements

The analyses can be run on a standard desktop computer. The replicator should expect the code to run for about ~2 hours.

Folders and Contents

A. Data

The Data folder contains the following datasets, organized in subfolders according to the underlying source of data:

- Subfolder "bls"
  - "isco_soc_crosswalk_stata.xlsx": The ISCO-08 / 2010 SOC crosswalk produced by the U.S. Bureau of Labor Statistics. Retrieved 05/26/21, from https://www.bls.gov/soc/isco_soc_crosswalk.xls. Included.
  - "national_M2018_dl.xlsx": The 2018 Occupational Employment and Wage Statistics database at the national level produced by the U.S. Bureau of Labor Statistics. Retrieved 06/04/21 from https://www.bls.gov/oes/special-requests/oesm18nat.zip. Included.
- Subfolder "cait"
  - "CW_CAIT_GHG_Emissions.xlsx": Emission Series produced by Climate Watch, disaggregated by country, emitting sector, and gas type. Retrieved 10/06/21, from https://www.climatewatchdata.org/data-explorer/, selecting "Download Bulk Data" option, and then "GHG EMISSIONS". Included.
- Subfolder "hci"
  - "HCIEXCEL.xlsx": The 2020 World Bank's Human Capital Index database in Excel format. Retrieved 03/16/21, from https://datacatalog.worldbank.org/search/dataset/0038030. Included.
- Subfolder "ilo"
  - "EMP_NIFL_SEX_RT_A-full-2021-12-27.dta": A country-year panel database with the "Informal employment rate by sex (%)" indicator. Retrieved 12/27/21, from https://rshiny.ilo.org/dataexplorer10/?lang=en, selecting "Informal employment rate by sex (%) – Annual", and then clicking on "Export" and selecting "full" and "dta" format. Included.
  - "employment_sex_2d.dta": A country-year panel database with the "Employment by sex and occupation - ISCO level 2 (thousands)" indicator. Retrieved 06/25/21, from https://rshiny.ilo.org/dataexplorer10/?lang=en, selecting "Employment by sex and occupation - ISCO level 2 (thousands) – Annual", and then clicking on "Export" and selecting "full" and "dta" format. Included.

- Subfolder “oecd”
  - “environmentally_related_tax_revenue.xlsx”: A country-year panel database with the “Tax revenue, % of GDP, All tax bases, Total environment” indicator. Retrieved 08/03/21, from https://stats.oecd.org/. In the left panel, select “Environment”, then “Environmental Policy”, then “Environmentally related tax revenue”, and then “All countries”. Customize the database according to the desired structure and download. Included.
  - “environmental_policy_stringency.xlsx”: A country-year panel database with the “Environmental Policy Stringency” indicator. Retrieved 08/03/21, from https://stats.oecd.org/. In the left panel, select “Environment”, then “Environmental Policy”, then “Environmental Policy Stringency”, and then “Environmental Policy Stringency Index”. Customize the database according to the desired structure and download. Included.
  - “green_growth_indicators.xlsx”: A country-year panel database with several green growth indicators. Retrieved 06/30/21, from https://stats.oecd.org/. In the left panel, select “Environment”, then “Green Growth”, then “Green Growth Indicators”, and then “All indicators”. Customize the database according to the desired structure and download. Included.
  - “trade_in_environmentally_related_goods.csv”: A country-year panel database with “Trade in Environmentally Related Goods” indicators. Retrieved 09/03/21, from https://stats.oecd.org/. In the left panel, select “Environment”, then “Policy Indicators on Trade and Environment”, and then “Trade in Environmentally Related Goods”. Customize the database according to the desired structure and download. Included.
- Subfolder “o-net”
  - “2010_Occupations.xlsx”: A list of all O*NET-SOC 2010 codes, titles, and descriptions. Retrieved 06/03/21, from the O*NET Resource Center (https://www.onetcenter.org/dl_files/database/db_28_2_excel/Occupation%20Data.xlsx). Included.
  - “Green Occupations.xlsx”: A list of all green occupations with their corresponding category, O*NET-SOC Code and title. Retrieved 05/18/21, from the O*NET Resource Center (https://www.onetcenter.org/dl_files/database/db_22_0_excel/Green%20Occupations.xlsx). Included.
  - “Related Occupations Matrix.xlsx”: The O*NET matrix of related occupations with career changers and career starters used to define green rival occupations. Retrieved 06/11/21, from the O*NET Resource Center (https://www.onetcenter.org/dl_files/database/db_28_2_excel/Related%20Occupations.xlsx). Included.
- Subfolder “own production”
  - This is the subfolder containing all databases produced in the project. These databases are generated by executing the do-files below, utilizing the other databases in the "/Data" folder as inputs. Included.

- Subfolder “sedlac”

  - This subfolder contains all the restricted-use databases generated in the project and sourced from the SEDLAC harmonization project (World Bank and CEDLAS). The original data source is the World Bank Stata API *datalibweb,* for internal use only. See the “Data Availability Statement” below in this document for instructions to access de SEDLAC project data using *dataliweb*.

- Subfolder “wdi”

  - “CLASS.xlsx”: The World Bank Group country classifications by income level. Retrieved 10/12/21, from https://datacatalogfiles.worldbank.org/ddh-published/0037712/DR0090755/CLASS.xlsx. Included.

  - “controls.xlsx”: A country-year panel database with several indicators on education and employment. Retrieved 09/02/21, from World Development Indicators - WDI (https://databank.worldbank.org/source/world-development-indicators). Select All countries, the series in the “Series – Metadata” excel sheet and All years. Then customize in the “Layout” section with series in columns, and country and time in rows. Included.

  - “environmental_variables.xlsx”: A country-year panel database with several indicators on education and employment. Retrieved 08/20/21, from World Development Indicators – WDI (https://databank.worldbank.org/source/world-development-indicators). Select All countries, the series in the “Series – Metadata” excel sheet and All years. Then customize in the “Layout” section with series in columns, and country and time in rows. Included.

  - “gdp.xlsx”: A country-year panel database with different measures of GDP. Retrieved 09/03/21, from World Development Indicators – WDI (https://databank.worldbank.org/source/world-development-indicators). Select All countries, the series in the “Series – Metadata” excel sheet and All years. Then customize in the “Layout” section with series in columns, and country and time in rows. Included.

  - “gdp_pc.xlsx”: A country-year panel database with GDP per capita, added value, and labor market indicators. Retrieved 08/02/21, from World Development Indicators – WDI (https://databank.worldbank.org/source/world-development-indicators). Select All countries, the series in the “Series – Metadata” excel sheet and All years. Then customize in the “Layout” section with series in columns, and country and time in rows. Included.

B. Do_files

Contains Stata .do files that replicate the data analysis (main paper and appendix).

“0_master.do”: **To run all the analysis, open this do file and set the path in line 29 to point to the folder on your computer** that contains this archive. Running this .do file calls all the .do files needed to create the tables and figures in the paper, which are the following:

- **“01_GHGE_data.do”:** Selects sectors and years from the CAIT GHGE database to be used for

constructing the new estimates and reclassifies sectors according to the harmonized classification CAIT-OECD-SEDLAC.

- **"02_GHGE_estimation_world.do":** Estimates the new GHGE using OCDE OITs for all countries, generates technology coefficients, and sectors shares in total emissions.
- **"03_GHGE_oecd_lac.do": S**elects sectors and years from the OECD Stats GHG emissions database to be compared with the new estimates.
- **"04_estimates_comparison_LAC.do":** Compares the estimates with OCDE, CAIT, and WDI GHG emissions data. This do-file produces Table 6 in the paper.
- **"05_GHGE_estimation_LAC.do":** Estimates the new GHGE using OCDE OITs for LAC countries only, generates technology coefficients, sectors' shares in total emissions, and final demand vectors.
- **"06_emissions_per_worker.do":** Estimates the number of emissions per worker and classifies the economic sectors into green and not green according to their level of emissions per worker. This do-file produces Table 7 in the paper.
- **"07_green_occ.do"**: Generate databases with green occupations at the different digit levels of the ISCO-08.
- **"08_process_ilostat.do"**: Process employment database from ILOSTAT and merges it with the green occupations at the 2-digit level of the ISCO-08 generated in "1_green_occ.do" to estimate the share of green jobs by country and year. This do-file produces figures 2 and 3 in the paper, and A2 in the appendix.
- **"09_merges_ilo_sedlac.do"**: Merge the green share database from ILOSTAT obtained in "2_process_ilostat.do" with a similar green share database obtained from SEDLAC.
- **"10_cross_country_stats.do"**: Generate cross-country statistics based on the last available data in the ILO/SEDLAC green share database generated in "3_merges_ilo_sedlac.do" and WDI data. This do-file produces Figures 4 and 15 in the paper, and Figure A6 and Table A1 in the appendix.
- **"11_lac_appends.do":** Creates LAC databases using SEDLAC to be merged with emissions data and green jobs data.
- **"12_green_jobs_circa_2019.do":** Calculates estimates of green jobs for the LAC region and by country, using circa 2019. This do-file produces Figures 9 to 14 in the paper, and Table A7 in the appendix.
- **"13_non_green_jobs_agriculture.do":** Estimates the share of non-green jobs in agricultural employment using circa 2019. This do-file produces Figure 6 in the paper, and Table A8 in the appendix.
- **"14_green_jobs_lac_series.do":** Creates the green job series of the LAC region. This do-file produces Figure 7 in the paper.
- **"15_regressions.do":** Estimates various sets of green jobs regressions using circa 2019. This do-file produces Table 8 in the paper.
- **"16_process_environmental_variables.do"**: Process OECD and WDI databases and merge them to the green share database from ILO/SEDLAC.
- **"17_environmental_regressions.do"**: Runs regressions of green employment on environmental variables. This do-file produces Table A4 in the appendix.
- **"18_descriptive_sedlac.do"**: Generate descriptive statistics of green employment based on SEDLAC. This do-file produces Table 9 in the paper.
- **"19_decompo_emissions.do"**: Decomposes the change in emissions over time. This do-file produces Figure 8 in the paper.

- **“20_green_jobs_circa_2019_Annex.do”:** Calculates estimates of green jobs for the LAC region and by country, using circa 2019 and a different sectoral classification. This do-file produces Figures A3 and A5 in the appendix.

C. Outputs

The .do files export tables in Excel format to the folder *Outputs*. This folder contains .xls and .xlsx files reproducing all the tables in the paper created as part of the data analysis, and the raw data underlying all figures that are not conceptual. Notice that the name of the exported output files matches the name of the corresponding table or figure. The tables below enumerate all the tables and figures in the paper with the program and line that reproduces them. Conceptual Tables and Figures not resulting from data analysis are marked as “Not empirical”.

C.1. Tables

Table #	Program	Line number in the program
Main Tables		
1	Not empirical	
2	Not empirical	
3	Not empirical	
4	Not empirical	
5	Not empirical	
6	04_estimates_comparison_LAC.do	58
7	06_emissions_per_worker.do	205
8	15_regressions.do	135-151
9	18_descriptive_sedlac.do	99-131
Appendix		
A1	10_cross_country_stats.do	124
A4	17_environmental_regressions.do	57-67
A7	12_green_jobs_circa_2019.do	176-180
A8	13_non_green_jobs_agriculture.do	163

C.2. Figures

Figure #	Program	Line Number in the program
Main Figures		
1	Not empirical	
2	08_process_ilostat.do	165
3	08_process_ilostat.do	168
4	10_cross_country_stats.do	107
5	12_green_jobs_circa_2019.do	295
6	13_non_green_jobs_agriculture.do	137
7	14_green_jobs_lac_series.do	151
8	19_decompo_emissions.do	192
9	12_green_jobs_circa_2019.do	330
10	12_green_jobs_circa_2019.do	355
11	12_green_jobs_circa_2019.do	380

Figure #	Program	Line Number in the program
Main Figures		
12	12_green_jobs_circa_2019.do	410
13	12_green_jobs_circa_2019.do	437
14	12_green_jobs_circa_2019.do	464
15	10_cross_country_stats.do	116
Appendix		
A2	08_process_ilostat.do	170-179
A3	20_green_jobs_circa_2019_Annex.	175
A5	20_green_jobs_circa_2019_	203
A6	10_cross_country_stats.do	283-285

References

Di Maro, V., K. Montoya, S. Olivieri, E. Vazquez and H. Winkler (2024). “Measuring Green Jobs: A New Database for Latin America and Other Regions.” World Bank Policy Research Working Paper.

Data Availability Statement

"Measuring Green Jobs: A New Database for Latin America and Other Regions"

The data used in the study is publicly available and can be found in the reproducibility package, except for the household surveys microdata from the Socio-Economic Database for Latin America and the Caribbean (CEDLAS and The World Bank), which are only available to members of the World Bank Group.

Datalibweb access to data

DATALIBWEB is a data system specifically designed to enable users to access the most up to date versions of non-harmonized (original/raw) and harmonized datasets of different collections across Global Practices. It can perform computations relevant for poverty and shared prosperity analysis based on the microdata from various harmonized collections: EAPPOV, ECAPOV, MNAPOV, SARMD, SEDLAC, LABLAC, SSAPOV, and the global collection GPWG/GMD. It also can help users to have access to the original/raw data for further country-specific analysis.

There is both the STATA program <datalibweb> as well as the website with the furl “datalibweb/” for subscribing.

Each person that wants to download data in Stata using the program code <datalibweb> needs to subscribe personally within the “datalibweb/” website, and sign the General Disclaimer on Terms of Use (see below). To assure that users have the most up to date microdata, and to make sure outputs/projects can be replicated by the original user or others, we highly recommend that people do not download data onto their computers, but rather form the habit of accessing it using the <datalibweb> program in STATA (incorporating it into their do-file codes), as explained further below. Users should always note the vintage of the data.

If you have not installed DATALIBWEB, please follow these instructions:

1. type in the intranet "datalibweb/",
2. go to the tab “about”,
3. follow the instructions below “manual installation”:
    a. the “automatic” option is not working, so please follow the instructions related to “manual installation”,
    b. follow the steps exactly. If you change the path, or modify the instructions, <datalibweb> may not work,
    c. restart your pc after manual installation of the package,
    d. if you have problems related to the installation, please email the <datalibweb> team at: datalibweb@worldbank.org

After <datalibweb> is installed, you must:

1. accept the General Disclaimer in the datalibweb/ website (the red box in the upper right corner needs to turn green)
2. check if the data you want to access is public or private,
    a. if public: you have access immediately via STATA,

    b. if private: you must request access, with an explanation for the use of the data, and the TTL will have to approve (the TTL has the MOU with the country NSO in these cases),
3. please note you can search for SEDLAC or LABLAC data under “Select Server/Country”,

In STATA:

1. Once you receive access, you can then use <datalibweb> within STATA to call the microdata.
2. Within STATA, type “help datalibweb” to see examples of the syntax/program code that loads the microdata into STATA.
    a. Example: datalibweb, country(ecu) years(2017) type(SEDLAC-03) mod(pov) clear
    b. This code opens the POVerty module for Ecuador 2017 from the version 03 of the SEDLAC project.