

README

Data and Code for “Occupational Mobility and Green Transition: A Stylized Estimation of Skill Investment Needs in South Asia”

by Andrés Ham, Emmanuel Vazquez, Monica Yanez-Pagans,
Camilla Knudsen and Saher Asad¹
(Mar 2026)

Overview

This document provides information needed to replicate the results in “Occupational Mobility and Green Transition: A Stylized Estimation of Skill Investment Needs in South Asia”.

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 Ham is with the Department of Economics and the School of Government, Universidad de los Andes. Vazquez is with CEDLAS (IIE, FCE), Universidad Nacional de La Plata. Yanez-Pagans, Knudsen, and Asad are with the World Bank. Corresponding author for the reproducibility package: Emmanuel Vazquez (evazquez@cedlas.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 ~5 minutes.

Folders and Contents

A. Data

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

- Subfolder “bls”

- o “isco_soc_crosswalk_stata.xlsx”: The ISCO-08 / 2010 SOC crosswalk produced by the U.S. Bureau of Labor Statistics. Retrieved on 05/26/21 from https://www.bls.gov/soc/isco_soc_crosswalk.xls. Included.

- o “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 on 06/04/21 from <https://www.bls.gov/oes/special-requests/oesm18nat.zip>. Included.

- Subfolder “esco”

- o “greenSkillsCollection_en.csv”: A dataset with green skills from ESCO version 1.2.0. Retrieved on 05/27/2024 from <https://esco.ec.europa.eu/en/use-esco/download>. Included.

- o “occupationSkillRelations_en.csv”: A dataset linking skills to occupations from ESCO version 1.2.0. Retrieved on 05/27/2024 from <https://esco.ec.europa.eu/en/use-esco/download>. Included.

- o “occupations_en.csv”: A dataset linking ESCO occupations to ISCO-08 from ESCO version 1.2.0. Retrieved on 05/27/2024 from <https://esco.ec.europa.eu/en/use-esco/download>. Included.

- Subfolder “granata and posadas”

- o “Green occupations.xlsx”: A dataset with the Green Task Intensity for each ISCO-08 code at the 4-digit level, from the toolkit in Granata and Posadas (2024), “Why Look at Tasks When Designing Skills Policy for the Green Transition? A Methodological Note on How to Identify Green Occupations and the Skills They Require”. Available at <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099507304172419760>. Shared by email by Josefina Posadas on 10/19/2022. Included.

- Subfolder “sar household surveys”

- o “BGD 2022 QLFS V01 M V01 A GLD ALL.dta”: Bangladesh – Quarterly Labour Force Survey 2022. Harmonized microdata from the Global Labor Database (GLD), World Bank.

- o "LFS2022_PUF.dta": Bhutan – Labour Force Survey 2022. Shared by the government with the World Bank Group.
- o "IND_2022_PLFS_V01_M_V02_A_GLD_ALL.dta": India – Periodic Labour Force Survey 2022/23. Harmonized microdata from the Global Labor Database (GLD), World Bank.
- o "hies_2019.dta": Maldives – Household Income and Expenditure Survey 2019. Shared by the government with the World Bank Group.
- o "NLSS23_raw.dta": Nepal – Living Standards Survey 2022/23. Shared by the government with the World Bank Group.
- o "PAK_2020_LFS_V01_M_V04_A_GLD_ALL.dta": Pakistan – Labour Force Survey 2020/21. Harmonized microdata from the Global Labor Database (GLD), World Bank.
- o "LKA_2021_LFS_V01_M_V03_A_GLD_ALL.dta": Sri Lanka – Labour Force Survey 2021. Harmonized microdata from the Global Labor Database (GLD), World Bank.
- Subfolder "ISCO correspondence"
 - o "corrtab08-88.xls": The ISCO-08 to ISCO-88 correspondence table produced by the International Labour Organization (ILO). Retrieved on 12/07/22 from <https://ilostat.ilo.org/methods/concepts-and-definitions/classification-occupation/>. Included.
 - o "ISCO.xlsx": The "Excel file with ISCO-08, -88, -68, -58 and skill levels" produced by ILOSTAT. Retrieved on 04/24/23 from <https://ilostat.ilo.org/methods/concepts-and-definitions/classification-occupation/>. Included.
- Subfolder "o-net"
 - o "Green Occupations.xlsx": A list of all green occupations with their corresponding categories, O*NET-SOC code and title. Retrieved on 05/18/21 from the O*NET Resource Center (https://www.onetcenter.org/dl_files/database/db_22_0_excel/Green%20Occupations.xlsx). Included.
 - o "Related Occupations Matrix.xlsx": The O*NET matrix of related occupations with career changers and career starters used to define green rival upskilling and reskilling occupations. Retrieved on 06/11/21 from the O*NET Resource Center (https://www.onetcenter.org/dl_files/database/db_28_2_excel/Related%20Occupations.xlsx). Included.
 - o "2010 Occupations.xlsx": A list of all O*NET-SOC 2010 codes, titles, and descriptions. Retrieved on 06/03/21 from the O*NET Resource Center (https://www.onetcenter.org/dl_files/database/db_28_2_excel/Occupation%20Data.xlsx). Included.
- Subfolder "own production"
 - o 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.

- Subfolder “skill council for green jobs”

- o “Qualifications database - update 2024.xlsx”: A list of green job qualifications and associated occupations based on ISCO-08 from the India Skill Council for Green Jobs, presented in Table A.7. Compiled on 02/01/2025 from Skill Council for Green Jobs (2024), “Green Jobs Handbook”, as well as qualification packs from the SCGJ webpage (<https://sscgi.in/>) and qualification files and model curricula from the National Qualifications Register (NQR) webpage (<https://www.nqr.gov.in/>). 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 28 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:

- “1_green_def_iscos.do”: Generate datasets with green and non-green occupational categories based on O*NET at different ISCO-08 digit levels.
- “2_gti_iscos.do”: Generate datasets with the Green Task Intensity (GTI) at different ISCO-08 digit levels.
- “3_scgj_iscos.do”: Generate datasets with the India SCGJ green definitions at different ISCO-08 digit levels.
- “4_esco_iscos.do”: Generate datasets with the ESCO green definitions at different ISCO-08 digit levels.
- “5_merge_definitions.do”: Merge green definitions across different ISCO-08 digit levels.
- “6_process_hhsurveys.do”: Process household surveys in South Asian countries, merging them with the corresponding green job definitions datasets.
- “7_descriptives.do”: Obtain descriptive statistics based on the processed household surveys in in South Asian countries.

C. Outputs

The .do files export tables in Excel format to the *Outputs* folder. This folder contains .xls files reproducing all tables in the paper that were generated as part of the data analysis, as well as the raw data underlying all figures. The name of each exported file matches the corresponding table or figure.

Some tables in the paper do not present analysis results. These include conceptual tables describing assumptions or definitions used in the analysis. For these items, no replication code is provided.

The tables below list all tables and figures in the paper and indicate, for each item, whether it is reproduced by code and, when applicable, the program and line that generate it.

C.1. Tables

Table #	Program	Lines
Main Tables		
1	Does not show analysis results	
2	Does not show analysis results	

3	7_descriptives.do	22-60
4	7_descriptives.do	66-133
5	7_descriptives.do	66-133
6	7_descriptives.do	273-339
Appendix		
A1	Does not show analysis results	
A2	5_merge_definitions.do	40-41
A3	5_merge_definitions.do	44-45
A4	5_merge_definitions.do	48-49
A5	5_merge_definitions.do	52-53
A6	5_merge_definitions.do	56-57
A7	Does not show analysis results	
A8	5_merge_definitions.do	60-61
A9	5_merge_definitions.do	64-65
A10	7_descriptives.do	215-217
A11	7_descriptives.do	223-225
A12	7_descriptives.do	227-240
A13 *	7_descriptives.do	345-415
A14	7_descriptives.do	248-250

C.2. Figures

Figure #	Program	Lines
Main Figures		
1	7_descriptives.do	66-139
2	7_descriptives.do	66-139
3	7_descriptives.do	66-139
4	7_descriptives.do	66-139
5	7_descriptives.do	145-192
6	7_descriptives.do	198-261
7	7_descriptives.do	198-261

* In the particular case of Table A13, the output includes columns with standard errors that must be hidden but used to assign the corresponding number of significance stars, based on whether the absolute value of the difference divided by the corresponding standard error exceeds the usual critical values of the t-test. In Excel, this can be done using the formula =IF(ABS(B4/C4)>2.57583;CONCATENATE(ROUND(B4*100;1);"%***");IF(ABS(B4/C4)>1.95996;CONCATENATE(ROUND(B4*100;1);"%**");IF(ABS(B4/C4)>1.64485;CONCATENATE(ROUND(B4*100;1);"%*");CONCATENATE(ROUND(B4*100;1);"%")))), where B4 and C4 are, respectively, the difference in sample means and the standard error of the difference for a particular characteristic in a given country.

References

Ham, A., E. Vazquez, M. Yanez-Pagans, K. Knudsen, and S. Asad (2026). "Occupational Mobility and Green Transition: A Stylized Estimation of Skill Investment Needs in South Asia". World Bank Policy Research Working Paper.

Data Availability Statement

" Occupational Mobility and Green Transition: A Stylized Estimation of Skill Investment Needs
in South Asia "

The data used in this study are publicly available and can be found in the reproducibility package, except for the household survey microdata, which come from the Global Labor Database (GLD) or were privately shared by country governments with the World Bank. These data are only available to members of the World Bank Group.