

This document explains the steps necessary to replicate the data, figures, and tables used in Belhaj Hassine Belghith N., Fernandez F.C., Lavin, B.A. (2025). Inequality, Education, and Occupational Change in the Philippines. WB PRWP 11163

This repository includes the material necessary to replicate the study's findings. The folders include the harmonized data used for the analysis, the codes and the outputs.

Original data can be requested on the Philippine Statistic Authority Website:

<https://psada.psa.gov.ph/catalog/LFS/about>. Users can register on the website and download the raw data files for each year from “LFS_PUF_October_2002” to LFS_PUF_October_2022”, LFS_PUF_July_2023”, and “LFS PUF January 2024”. Data will be in CSV and DAT formats and users need to download the data dictionary in xlsx format for each data file.

The analysis uses STATA software. The codes need to be run in two parts:

- 1) the codes for the harmonization of the data: change the directory for main1.do in the “0-Harmonization\Do Files Harmonization LFS” folder to run the code for harmonization. This will export the harmonized data in “All years” folder. Copy this to “2-data” folder to run the next step.
- 2) the codes for the analysis in the paper: change the directory for main2.do in the “1-code” folder.

The codes for the harmonization of the data convert the CSV files into Stata files (oct_2022 to oct_2022.dta, jul_2023.dta and jan_2024.dta), harmonize and append the data. Data is harmonized to create the occupation, education and wage variables as explained below. Final harmonized data for each year are appended together and restricted to the sample of workers used in the analysis as explained below.

Seven Stata data files were created to harmonize the variables over time (EducLabel_2019.dta
IndLabel_2012_2013.dta IndLabel_2016_2019.dta OccLabel_2020.dta
PSIC_2_to_1_digit.dta PSOC_1992to2012.dta PSOC_2012balancedocc.dta)
These data files will be uploaded to Data Development Hub.

The codes to execute the analysis in the paper include 4 do files:

1. 1-Wage premium_occupation.do estimates the composition adjusted wages, hours worked, skill premium and skill supply. It produces figures 1-7, 9-11 and figures 1A and 2A in the appendix.
2. 2-RIF regression>Returns to education & occupation.do runs the unconditional quantile RIF regression and produces figure 8, figure 3A and tables 1A and 2A in the appendix.
3. 3-Reweight wage change_DFL estimates the reweighted densities and counterfactual analysis of wage changes across occupations. It produces figure 12 and 4A in the appendix.
4. 4- Employment Statistics_LFS generates employment and unemployment descriptive statistics in figures 13 and 5A.

The do files (1-3) are based on the codes of Nicole Fortin (2018) and David Autor (2011, 2019). Please note that income Gini coefficients in Figure 1 are from FIES and can be found in WDI (<https://databank.worldbank.org/source/world-development-indicators>). As indicated in footnote 5

in the paper, as income Gini data from FIES is available only every three years (and every two years since 2021), we assumed constant values between survey years.

Data Statement

This research uses harmonized data that were constructed from the raw data available on the PSA OpenSTAT Portal, which is accessible upon registration through the PSA website. While the harmonized datasets used in this paper can be made available, any republication or resharing must comply with the guidelines outlined in the PSA access conditions: <https://psada.psa.gov.ph/access-conditions>.

Instructions for Replicators

Users must first have access to all data files if they are not included in the reproducibility package. They should request the data, download the listed files, and place them in the data folder. - Update the do files with your directory paths

- The do files for the analysis can be executed separately to produce the needed outputs
- Ensure all required STATA software ado files are installed (i.e., fastgini, ineqdeco, pctile etc.). Use Stata ssc install package_name.

Harmonization of Occupations

For this paper, it was necessary to create an original crosswalk to categorize occupations as the Philippines uses its own national system, The Philippine Standard Occupational Classification (PSOC), which had not been applied to groupings common in this literature. The Labor Force Survey (LFS) rounds from January 2002 to January 2016 uses the 1992 PSOC which is modeled after the International Standard Classification of Occupations 1988 (ISCO-88). Starting from April 2016 and continuing through end of the period covered by this paper, the LFS uses the 2012 PSOC, which based on ISCO 2008 (ISCO-08).

The 1992 PSOC is comprised of 31 occupational categories at the two-digit level (excluding Armed Forces occupations) with 9 major groups at the one-digit level. These groups are: (1) Managers, (2) Professionals, (3) Technicians and Associate Professionals, (4) Clerical Support Workers, (5) Service and Sales Workers, (6) Skilled Agricultural, Forestry and Fishery Workers, (7) Craft and Related Trades Workers, (8) Plant and Machine Operators and Assemblers, and (9) Elementary Occupations. The 2012 PSOC disaggregates multiple occupational categories from the 1992 PSOC, resulting in 40 occupational categories that are organized under 9 major groups that closely mirror those in the 1992 PSOC. The revision introduced new categories aimed at increasing granularity: for example, the major group *Professionals*, previously with 4 categories, was expanded to 6 categories; *Laborers and unskilled workers*, initially with 3 categories, was further disaggregated to 6 categories.

To create continuity across the two occupational classifications, a set of harmonized occupations was constructed at the two-digit level, with 22 distinct occupational categories. These harmonized

occupational classifications were then grouped into five broader categories of occupation modeled after Dorn (2009), Acemoglu and Autor (2011) and Bárány and Siegel (2018). Specifically:

1. High-skill occupations comprising of managers and managing proprietors, professionals, and associate professionals and technicians
2. Middle-skill routine occupations including clerical support workers, sales workers, craft and related trades workers, plant and machine operators and assemblers
3. Middle-skill nonroutine occupations include personal and protective service workers
4. Low-skill occupations encompass elementary occupations, such as cleaners and helpers as well as laborers in mining, construction, manufacturing and transport

We focus our analysis on a sample of wage workers aged 15 and above. The analysis therefore excludes the following categories of workers which the Philippine Statistics Authority (PSA) classifies as non-wage workers: 1) self-employed without any paid employee, 2) employer in own family-operated farm or business, and 3) worked with or without pay in own family-operated farm or business. Following Bárány and Siegel (2018), workers with missing wage data and those in armed forces occupations were excluded from the analysis.

Harmonization of Educational Attainment

The period from 2002 to 2024 covers three rounds of the Philippine Standard Classification of Education (PSCED). Specifically, October LFS rounds from 2002 to 2015 utilize the 1994/1997 PSCED, patterned after the 1976 International Standard Classification of Education (1976 ISCED), rounds from 2016 to 2018 utilize the 2008 PSCED, patterned after the 1997 ISCED, while beginning in 2019, October LFS rounds have utilized the 2017 PSCED, which is aligned with both the 2011 ISCED as well as the 2013 ISCED Fields of Education (ISCED-F).

The transition across PSCED rounds reflect important changes that occurred in the basic education system as a result of the K-to-12 transition, which lengthened compulsory education from 10 years to a total of 13 years, encompassing kindergarten *and* 12 years of basic education. LFS rounds beginning in 2016 reflected these changes, with PSCED categories for educational attainment revised to both align with the new system while maintaining consistency with previous PSCED rounds.

Educational levels were grouped into eight broad educational categories, encompassing the following: 1) No schooling, 2) Incomplete elementary, 3) Elementary Graduate, 4) Incomplete high school, 5) High school graduate, 6) Post-secondary/TVET, 7) Incomplete college and 8) College graduate and above. To ensure consistency and comparability across rounds, particularly during years when the transition occurred, education statistics were generated and benchmarked against figures published by the PSA.

Construction of Wage Variable

Real weekly wages were used for the analysis to ensure comparability across years. These were calculated based on the daily wage and the numbers of hours worked per week, as reported in the LFS. The analysis excludes the following categories of workers which the PSA classifies as non-wage workers: 1) self-employed without any paid employee, 2) employer in own family-operated

farm or business, and 3) worked with or without pay in own family-operated farm or business. Workers in salaried roles with missing data as well as workers paid on a commission basis or on a per piece basis were similarly excluded from the analysis. Wage are reported in real Philippine Pesos (PHP), deflated using the Philippine CPI.