

README

Data and Code for “Demographic Transition and Education Expenditure in South Asia: Opportunities and Challenges”

by Mariana Marchionni, Namrata Raman Tognatta, Emmanuel Vazquez and Monica Yanez-Pagans¹
(Jan 2026)

Overview

This document provides information needed to replicate the results in “Demographic Transition and Education Expenditure in South Asia: Opportunities and Challenges”.

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 Marchionni is with CEDLAS (IIE, FCE) – Universidad Nacional de La Plata and CONICET. Tognatta and Yanez-Pagans are with the World Bank, and Vazquez is with CEDLAS (IIE, FCE) – Universidad Nacional de La Plata. 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 ~8 minutes.

Folders and Contents

A. Data

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

- Subfolder “wb”
 - o “CLASS 2022 2023.xlsx”: The World Bank income classification of countries for the fiscal year 2022-2023. Retrieved on 04/28/23 from <https://blogs.worldbank.org/en/opendata/new-world-bank-country-classifications-income-level-2022-2023>.
- Subfolder “wpp”
 - o “WPP2022 POP F01 1 POPULATION SINGLE AGE BOTH SEXES.xlsx”: The 2022 *World Population Prospects* estimates and projections for both sexes and single ages. Retrieved on 05/22/23 from <https://population.un.org/wpp/>. Available by selecting “Data/Download Center”, then “Archive” under “File Type”, and clicking on “[WPP2022-Excel-files.zip \(ZIP\)](#)”.
 - o “school age population.xlsx”: The 2022 *World Population Prospects* estimates and projections based on the medium variant for both sexes and single ages. Retrieved on 05/22/23 from <https://population.un.org/wpp/>. Available by selecting “Data/Download Center”, then “Archive” under “File Type”, and clicking on “[WPP2022-Excel-files.zip \(ZIP\)](#)”. The Excel file includes the first two worksheets from “WPP2022_POP_F01_1_POPULATION_SINGLE_AGE_BOTH_SEXES.xlsx” and an additional worksheet named “School age population”, which appends estimates for 2020 and 2021 from the first worksheet to projections for 2023–2100 from the second worksheet. It adds columns L to V, which report the size and share of the population at ages corresponding to attendance at each education level, using the official entrance and completion ages reported in Table 5. The underlying source for these ages is the UNESCO Institute for Statistics (UIS).
 - o “population projections.xlsx”: A cleaned database ready to be imported into Stata, containing columns C (country), K (year), and R to V (the proportion of the total population attending each education level) for the countries of interest. The data are drawn from the worksheet “School age population” in the file “*school age population.xlsx*”.
 - o “WPP for elasticities.xlsx”: An Excel file derived from the 2022 World Population Prospects (WPP) estimates and projections, prepared for use in the estimation of elasticities. The file contains two worksheets. The first worksheet, “Estimates”, reproduces the WPP 2022 estimates for the period 1950–2021 for both sexes and single ages, corresponding to the same source and underlying data described in “*school age population.xlsx*”. In addition, it includes columns L to V, which report the size and share of the population at ages corresponding to attendance at each education level, constructed using uniform age ranges

applied across all countries: ages 4–6 for pre-primary education, 7–12 for primary education, 13–18 for secondary education, and 19–22 for tertiary education. The second worksheet, “Base”, contains the same information as the “Estimates” worksheet, but with the initial rows removed so that the first row includes only variable names, making the dataset directly importable into Stata.

- Subfolder “uis”

- o “status_quo.xlsx”: A cleaned database ready to be imported into Stata, containing 2020 data—or the latest pre-pandemic year available—on public expenditure on education as a share of GDP (g), gross enrollment rates (m), and the share of students in private institutions (1-pu), extracted on 13 Jul 2023 at 17:13 UTC (GMT) from UIS.Stat (<https://databrowser.uis.unesco.org/>). The proportion of the school-age population (n) corresponds to the 2020 estimates obtained from the World Population Prospects 2022 revision using the school ages reported in Table 5 (see “population_projections.xlsx”). Gross enrollment rates (m) for pre-primary and tertiary education were extracted on 04/18/23 from the World Development Indicators (WDI) (<https://databank.worldbank.org/source/world-development-indicators>). In the database, pu denotes the share of students in public institutions and is computed as the complement of the UIS-reported share in private institutions. Expenditure per student as a share of GDP per capita—Case p = 1 (b_y)—is obtained as $g/(p \cdot m \cdot n)$ to ensure an exact decomposition. For the same reason, expenditure per student as a share of GDP per capita—Case p = pu (b_y2)—is obtained as b_y/pu . The gross enrollment rate for total basic education is obtained as the sum, across primary and secondary education, of the product of the gross enrollment rates and the corresponding total school-age population, divided by the sum of the school-age population across these two levels, according to data published by the World Population Prospects (WPP) and available in “School age population.xlsx”. The same procedure is used to obtain the share of students in public schools (pu) for total basic education. Public expenditure on education as a share of GDP for total basic education is obtained as the sum of the respective expenditures at the primary and secondary levels. See Table A2 in the paper for a reference to the specific years used in each case.

- o “EDULIT_DS_20112019122436006.csv”: A database extracted on 20 Nov 2019 from UIS.Stat (<https://databrowser.uis.unesco.org/>), containing all available country-year observations (1970–2019) for a set of education indicators disaggregated by level of education. The dataset includes measures of enrollment, coverage, public spending per student, and school-age population. Specifically, the selected indicators comprise: (i) gross enrollment ratios for pre-primary, primary, lower secondary, upper secondary, secondary, and tertiary education; (ii) net enrollment rates for pre-primary, primary, lower secondary, upper secondary, and secondary education; (iii) initial government funding per student by education level, expressed as a percentage of GDP per capita, in nominal terms (USD and PPP), and in constant prices (USD and PPP); (iv) total enrollment by education level (number of students); and (v) school-age population by education level (number of individuals). These indicators are subsequently harmonized, reshaped, and combined with GDP per capita and demographic data for use in the empirical analysis.

- o “DEMO_DS_07112019102518924.csv”: A database extracted on 07 Nov 2019 from UIS.Stat (<https://databrowser.uis.unesco.org/>), containing all available country-year observations (1970–2019) on macroeconomic indicators related to income levels and price conversion. The dataset includes measures of GDP per capita in nominal and real terms, as well as exchange rates and purchasing power parity (PPP) indicators. Specifically, the selected indicators comprise: (i) GDP per capita in current U.S. dollars, in current local currency units

(LCU), and in current international dollars (PPP); (ii) GDP per capita in constant international dollars (PPP, base year 2011); (iii) the GDP deflator; (iv) the PPP conversion factor for GDP; (v) the official exchange rate (LCU per U.S. dollar); and (vi) the price level ratio of the PPP conversion factor to the market exchange rate. These indicators are reshaped and merged with education and demographic data to support the empirical analysis.

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 27 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_demographic_projections.do”:** Shows the WPP projected demographic evolution in the share of the school-age population, 1950–2050.
- **“2_expenditure_projections.do”:** Projects public expenditure on education for South Asia under different scenarios, 2020-2050.
- **“3_elasticities_appendix.do”:** Estimates the output elasticity of per-student benefits and the elasticities of coverage and benefits with respect to demographic changes.

C. Outputs

The .do files export tables and figures to the folder *Outputs*. This folder contains .xls and .xlsx files reproducing all tables in the paper that are generated as part of the data analysis, as well as files reproducing figures in image format (primarily .tif) or, in specific cases, in spreadsheet format containing the underlying data. The name of each exported output file matches the corresponding table or figure. For figures that display multiple country groups, separate image files are generated for the countries shown on the left and on the right of the figure, respectively, and are combined to produce the final figure as shown in the paper. In these cases, file names include suffixes “_1” and “_2”. For Figure 2, separate files are generated by education level, with file names indicating the corresponding level.

Some tables and figures in the paper do not show analysis results. These include: (i) items that report data directly taken from cited sources or involve simple arithmetic operations following procedures explicitly described in table notes, and (ii) conceptual tables describing assumptions or definitions used in the analysis. For these items, no replication code is provided.

The tables below enumerate all figures and tables 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. Figures

Figure #	Program	Lines
Main Figures		
1	1_demographic_projections.do	104-132
2	2_expenditure_projections.do	73-101
3	2_expenditure_projections.do	125-144
4	2_expenditure_projections.do	178-197
5	2_expenditure_projections.do	233-252
6	2_expenditure_projections.do	285-304
7	2_expenditure_projections.do	340-360
8	2_expenditure_projections.do	406-426
9	2_expenditure_projections.do	634-668

C.2. Tables

Table #	Program	Lines
Main Tables		
1	Does not show analysis results	
2	Does not show analysis results	
3	Does not show analysis results	
4	Does not show analysis results	
5	Does not show analysis results	
6	Does not show analysis results	
7	Does not show analysis results	
8	2_expenditure_projections.do	511-529
9	2_expenditure_projections.do	531-549
10	2_expenditure_projections.do	551-569
11	Does not show analysis results	
12	Does not show analysis results	
13	2_expenditure_projections.do	571-589
14	Does not show analysis results	
15	Does not show analysis results	
16	2_expenditure_projections.do	591-609
Appendix		
A1	3_elasticities_appendix.do	236-241
A2	Does not show analysis results	
A3	Does not show analysis results	
B1	3_elasticities_appendix.do	247-257
B2	2_expenditure_projections.do	611-628

References

Marchionni, M., N. R. Tognatta, E. Vazquez and M. Yanez-Pagans. "Demographic Transition and Education Expenditure in South Asia: Opportunities and Challenges". World Bank Policy Research Working Paper.