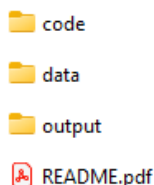


README

Data and Code for paper on intergenerational educational mobility in Bangladesh

1. Overview

This document contains all the information necessary to replicate the estimates presented in the paper *Converging Paths: Intergenerational Educational Mobility and the Decline of Gender and Geographic Gaps in Bangladesh*. The replication package is organized into the following folders and files:



2. Computational requirements

Full replication of the results requires the following software: Microsoft Excel, Stata, and Microsoft Word.

2.1 Stata

Stata version 14 or higher is required. The team produced this workflow using Stata 18. 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)

3. Memory and runtime requirements

The analysis can be run on any standard desktop or laptop computer without the need for specialized or high-performance computing infrastructure. A user should expect the complete execution of all code files, run in the correct sequence, to take less than minutes on a standard desktop computer, though actual runtime may vary slightly depending on the machine's specifications. The replication package has an initial file size of approximately 569 megabytes; however, users should ensure

sufficient additional storage space is available on their system, as the workflow will automatically generate a series of intermediate files during execution, increasing the total storage footprint considerably beyond the original size.

4. Folder and contents

The replication package is organized into two main folders. The **code** folder contains the Stata do-file (*IGM in education_BD.do*), which must be executed to reproduce all estimates presented in the paper. The **data** folder contains all the data files required for the package to run correctly, organized into two subfolders, which are described in detail below.

 harmonized files

 Rawdata_sec1_2

The `harmonized files` subfolder contains harmonized individual and income data files in Stata format (.dta), which are required for the code to run correctly. These files are derived from the Bangladesh Household Income and Expenditure Survey (HIES) 2022 microdata, provided by the Bangladesh Bureau of Statistics (BBS).

The HIES 2022 microdata is not publicly available and may not be redistributed. Users other than the original authors and designated replicators must request explicit access before using these files. Access can be obtained through the World Bank's internal data platform, *datalibweb*, or the SARMD (South Asia Regional Micro Database). To request access, users should submit a formal request via <http://datalibweb/> and receive approval from the Bangladesh team lead before the data can be loaded in Stata.

Once the user accesses or downloads the two harmonized files (BGD_2022_HIES_v01_M_v01_A_SARMD_INC.dta, and BGD_2022_HIES_v01_M_v01_A_SARMD_IND.dta), additional education-related variables are incorporated by further linking with the education module contained in the raw data files from the subfolder `Rawdata_sec1_2`, specifically the files `HH_SEC_2A.dta` and `HH_SEC_2B.dta`. These files capture the educational background of the parents of all household members, regardless of whether the parents themselves are listed in the household roster. The harmonized files, by contrast, only contain education information for individuals who appear directly in the household survey roster. The merge is therefore necessary to enrich the harmonized dataset with parental education information and is performed on the unique household and individual identifiers to ensure that records are correctly matched to their corresponding observations prior to any analytical work.

Since the paper was prepared as part of the Bangladesh Poverty and Equity Assessment (PEA), the harmonized data and the raw microdata used in the analysis were already available to the team and included in the data folder of the replication package. Users who already have approval for the data through *Datalibweb* and the Bangladesh raw HIES microdata through the Bangladesh team lead could therefore be able to proceed with replication without submitting an additional access request.

5. Do files

All analysis can be executed by running a single do-file (*IGM in education_BD.do*), which automatically generates all the tables and figures presented in the paper, with the exception of Figure 1, which is sourced from a separate report and is therefore not produced by this workflow.