Replication files for:

Preschool Availability and Female Labor Force Participation: Evidence from Indonesia

Daniel Halim, Hillary C. Johnson and Elizaveta Perova

Version: March 2021

This README file describes the contents of the `IndoPreschool` folder and instructs users to replicate tables and figures of the paper. All programs are written using `STATA` (version 15 or higher).

1. Folder system

- `Do files` folder contains all do-files to construct dataset and duplicate tables and figures
- `Results` folder is an empty folder to host generated tables and figures by `01a_master.do`, `02_tables.do`, and `03_figures.do`
- `Data` folder contains datasets to be used for replication.

2. Data

The paper draws from three main data sources.

- `Indonesian Family Life Survey (IFLS)`. Full access to all five waves of IFLS datasets are free to access upon registration at RAND website: https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS/access.html
- `PODES` is fielded roughly once every three years. We use all years between 1990 and 2014: 1990, 1993, 1996, 2000, 2003, 2005, 2008, 2011, 2014.
- `SUSENAS` is fielded once or twice per year depending on the year. We only use one round per year. In biannual survey years, we use June/July round. We only years that overlap with PODES years: 1990, 1993, 1996, 2000, 2003, 2005, 2008, 2011, 2014.

Full access to SUSENAS and PODES dataset are not publicly available, but can be purchased from Indonesia's Central Statistics Bureau (https://www.bps.go.id). For the purpose of replication, we include relevant variables (e.g. number of preschools and children aged 3-6) aggregated at the district level (circa 1993 district boundaries).

Belows are details of the dataset folder.

- `/Data/Raw data`, where raw IFLS, district codes crosswalk, shape files, and CPI data are located. List of folders
    - `IFLS`. The folder only contains a subset of IFLS datasets used in the analysis of this paper.
    - `District codes`: contains crosswalk datasets of districts codes from BPS across years.

    - `Shapefiles`: contains shape files of Indonesia by districts level for generating maps and spatial information of public preschool availability.
    - `nationalcpi.dta`, national consumer price index (CPI) with 2010 base year obtained from FRED. Accessible.
- `/Data/Clean data` contains mostly empty sub-folders, which will be the target file path for cleaned data processed by the cleaning do-files. Due to PODES and SUSENAS data restrictions, `/Data/Clean data/PODES` contains two cleaned data extracts to assist with replication purposes:
    - `kindergarten across podes.dta` contains the number of public and private preschools and the number of children aged 3-6 in PODES years.
    - `preschool density imputed.dta` contains the number of public and private preschools and the number of children aged 3-6 in all years (filled-in between PODES years with linear extrapolation method).

  It also contains the final cleaned data `preschool.dta`.

3. Do files:

Users will need to set up their working directory in `01a_master.do` by following these steps.

Setup global. To set the file path, modify global USER to 1. The global acts as a toggle to switch on the file path.

Example:

```
global USER                    1
```

Next, add your own global project by replacing `"<file path to replication folder>"` with the location of the Replication folder in your system

Example:

```
if $USER {

      global project ".../Documents/IndoPreschool"

      }
```

Please notice that file path subdirectory uses slash "/" instead of backslash "\". For Windows users, simply replace backslash "\" with slash "/" to set up your working directory.

Data cleaning

Users with full-access to the raw PODES and SUSENAS data (as described in Section 2) can immediately run `01a_master.do` to perform replication from data cleaning to produce tables and figures of the paper.

Alternatively, users without access full-access to the raw PODES and SUSENAS data should first modify `01a_master.do` by turning off the do-file cleaning PODES and SUSENAS data.

Example: adding an asterisk at the beginning of Line 87

```
*do "$dofile/01b_cleaning podes susenas.do" // to execute, need access to PODES
and SUSENAS dataset
```

Users can then run `01a_master.do` to perform replication, which will still clean the IFLS data from the raw data, but will call on clean PODES and SUSENAS data extracts instead.

The cleaning do-file(s) will ultimately produce the final cleaned data, `preschool.dta`, located at `/Data/Clean data`.

Data analysis

The analysis do-files will call on cleaned datasets to produce tables and figures in the paper, to be deposited at `/Results`.

- `preschool.dta` for reproducing Table 1, 2, 3, 4, 5, 6, 7, and Figures 2, 3. The cleaning do-files will ultimately produce this final cleaned data to be located at `/Data/Clean data`.
- `kindergarten across podes.dta`. for reproducing Figure 1 and summary statistics of public preschool availability at districts level in Appendix Table 1. Located at `/Data/Clean data/PODES`.

Users can skip the data cleaning do-files and jump straight to running the analysis do-files.

List of do files:

- `01a_master.do`, to perform full-process of replication from data cleaning to producing tables and figures in the paper. The do file also includes supplementary user-written ado files that are relevant to the replication.
- `01b_cleaning podes susenas.do`, to perform data cleaning, generating variables, and merging PODES and SUSENAS dataset. Executable upon access to the full PODES and SUSENAS raw datasets.
- `01c_cleaning ifls.do`, to perform cleaning, generating variables, and merging relevant IFLS datasets from all five waves. Users without access to the full PODES and SUSENAS raw datasets can skip `01b_cleaning podes susenas.do` and start the data cleaning process from here.
- `02_tables.do`, to produce Table 1, 2, 3, 4, 5, 6, 7 and Appendix Table 1. This do-file will deposit word documents (`.rtf`) containing tables as numbered in the paper.
- `03_figures.do`, to produce Figures 1 (Panel A and B), 2, and 3. This do-file will deposit `.eps` files containing figures as numbered in the paper.

4. Result files folder

This folder hosts all tables and figures including one appendix table.

5. Notes

Some redactional editing from the STATA output may be necessary to produce exact tables and figures in the paper. This editing only affects labels and display of some tables and figures, however shall not affect estimation results.