

Documentation file for Do-files and Datasets corresponding to paper titled:
“Turning a Shove into a Nudge? A "Labeled Cash Transfer" for Education”

by Najy Benhassine, Florencia Devoto, Esther Duflo, Pascaline Dupas and Victor Pouliquen

July 1, 2014

Files organization:

In order to generate the tables as in the paper, the analysis folder requires the following sub-folders:

- **“Dofiles”**: where all the analysis do-files should be placed.
- **“Input”**: where all the databases should be placed.
- **“Output”**: this folder will be filled with the intermediary databases generated by the do-files creating the variables needed for the analysis.
- **“Tables_paper”**: The folder where tables will be out-sheeted by the table creation code

Do Files

The main do-file is called **“lct_MASTER_analysis.do”**. It calls all the other do-files one by one.

The do-files merge all databases and generate the variables for analysis.

The last do-file called by the main do-file generates tables (**“lct_tables_creation.do”**).

Before running the **“lct_MASTER_analysis.do”** do-file you may need to change the path to the main analysis folder (Stata command: **“cd yourfiledirectory”**).

Data Sets

There are many databases needed to run the **“lct_MASTER_analysis.do”**. These databases come from 5 different data collection activities:

- Administrative databases from Tayssir administration:
 - **“cct_tayssir_admin_data_an.dta”**: Dataset with information on students enrolled in the Tayssir program, including number of days of absence by month and amount of transfer received by month.
- Household level data:
 - **“cct_baseline_an.dta”**: baseline household survey data
 - **“cct_endline_an.dta”**: endline household survey data
 - **“cct_monitoring_endline_an”**: Monitoring data with reason for survey non-response during endline household survey.
 - **“cct_hh_weights_an.dta”**: Database with sampling weights for household surveys.

- School level data:
 - **“cct_preliminary_survey_an”**: data from the school-level survey conducted in preparation for the study in 2008.
 - **“cct_school_visits_an.dta”**: Data from the 7 school visits done between 2008 and 2010.
 - **“cct_correction_dropout_date_an.dta”**: database created to input corrections to the school visit data about dropout dates. When dropout dates were not consistent between two visits, data were manually checked to determine which data was the most accurate.
 - **“cct_stratum_an.dta”**: database at the school sector level with the stratum used for the randomization.

- ASER data:
 - **“cct_aser_an.dta”**: data from tests in mathematics administered to one child per household during the endline household survey.

- Awareness surveys:
 - **“cct_knowledge_households_year1_an.dta”**: data from the awareness survey administered to a subset of households in 2009.
 - **“cct_knowledge_teachers_year1_an.dta”**: data from awareness survey administered to a subset of teachers in 2009.
 - **“cct_knowledge_teachers_year2_an.dta”**: data from awareness survey administered to subset of teachers in 2010.

All surveys were administered in French. All databases are labeled with the exact English translation of the questions.

All survey instruments used for the research are included in the “Survey_instruments” folder.

Codebooks:

For all databases, an Excel file describes all variables included in the datasets. Codebooks files are available in the “Codebook folder”. They are named with the prefix “codebook” followed by the name of the database it describes.