Readme file on Reproducibility: Direct and Indirect Effects of Subsidized Dual Apprenticeships

July 30, 2024

Overview

This document addresses the reproducibility of the paper "Direct and Indirect Effects of Subsidized Dual Apprenticeships". The following information addresses the aspects needed to run the code and data used for this publication.

What the package contains
The package folder, named **Reproducibility package** has the following items.

- The paper of the publication

- A folder, named *Raw datasets*, where the user will deposit the datasets used in this study, available in the World Bank Microdata Library and Enterprise Survey data portal.

- A folder, named *Paper 2nd follow-up* that contains the code and outputs produced for the paper.

- A folder named *Overleaf* that contains the latex file that compiles all figures and tables in Overleaf.

Computational requirements for the package

Paper exhibits were reproduced in a computer with the following specifications:

- OS: Windows 10 Enterprise, version 21H2

- Processor: Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz, 16 Core(s)

- Memory available: 32 GB

- Software version: Stata 18 Running the code takes approximately 5 hours runtime.

Stata packages used

The packages used for this publication are the following: *outtable, ietoolkit, estout, rwolf2, kdens, mhtreg, pdslasso, qqvalue, texdoc, touch, lasso2, ivqte*, and *outreg*. To prevent having any version control issues when reproducing results, the specific version used for each Stata package was saved in the *ado* folder. When running the code, please use the included versions rather than downloading them separately.

Instructions on running the code

Please follow the steps below to successfully reproduce the tables and figures of the paper.

1. Download the datasets used in the paper and save them to the *Raw datasets* folder. Follow the steps for each dataset in the data availability statement of this document.

2. Open the master Do-file (*Master Do-file.do*) in the *Paper 2nd follow-up* folder and set the path where you saved the reproducibility folder in line 37.

3. Run the master Do-file (it should be a one-button run) in Stata. The outputs are saved to four output sub-folders (details on this later) in the *Output* folder.

4. Take the four output subfolders (named *Graph*, *GraphOG*, *Tables*, *TableOT*) and upload them directly (as complete folders) to a new Overleaf session (in a personal Overleaf account you have).

5. Upload the *main.tex* file of the *Overleaf* folder and upload it to the Overleaf session where you uploaded the 4 output subfolders, and compile the file online to get the outputs seen in the publication paper.

The code in more detail

The master do-file does the following process:

- Does the initial set up

  - In line 37: sets the path of the reproducibility folder

  - In line 42: sets the *ado* folder with the Stata packages needed for the code to run.

- Runs the data processing given input raw and anonymized data. For this case, data cleaning and variable creation are both included in *processing do-file.do*.

- Runs *Outputs Restud paper.do*, which covers the tables and figures seen in the publication.

The data processing Do-file

processing do-file.do. cleans and produces variables from the raw datasets. It produces composite databases that contain baseline, midline, and endline variables for estimations.

The analysis Do-file

Outputs Restud paper.do covers the code used for the outputs (tables and figures) used in the publication. It first covers the 10 main tables of the paper[1], then covers all figures, and finally produces the 20 tables of the appendix.

Ouputs

The code produces all outputs of the paper, with the following exceptions (figures that do not need reproducibility verification given their nature):

- Figure 1: Experimental design.

- Figure 7: Youth decision to participate in apprenticeship.

- Figure 8: Equilibrium employment of traditional apprentices and tightness.

There are four subfolders of outputs[2]:

- *Tables*: main tables

- *TableOT*: appendix tables

- *Graph*: main figures

- *GraphOT*: appendix figures

Data availability statement

This publication uses four types of datasets that are publicly available in the World Bank Microdata Library (MDL) and the Enterprise Survey data portal. Below is the information on the name of the datasets, available documentation, and instructions for accessing each.

---

[1]The only exception is Table 1, which is produced after table A2. This because it uses some global macros from tables A1 and A2.

[2]As mentioned before, please upload these four folders (full folders, not outputs separately) to Overleaf, along with the main.tex file in the Overleaf folder and compile online.

Enterprise National Survey in Côte d'Ivoire for year 2016 [3]

- Is this dataset included in the reproducibility package: no.

- Dataset available in this MDL website.

- Accompanying documentation in the MDL website for this dataset: questionnaire and technical reports on its implementation.

- Instructions on accessing this dataset in the MDL: please click on the **get microdata** tab. Register in the WB Enterprise Data portal and click on a folder named *Côte d'Ivoire-2016-full data.dta*. In it you will find the Stata file called **C-te-d-Ivoire-2016-full-data.dta**

- Instructions for this file: Please download **C-te-d-Ivoire-2016-full-data.dta** and locate it at
  Raw datasets/Enterprise survey in the reproducibility folder.

Firm surveys: baseline and midline datasets[4]

- Are these datasets included in the reproducibility package: no.

- Datasets are available in this MDL website

- Instructions on accessing these datasets in the MDL website: click on the **get microdata** tab. Register in the data portal and download the folder named **Data in Stata14 Format**.

- After downloading, please locate the four .dta datasets in the **Raw datasets/Firms** subfolder.

Baseline firm survey dataset

- Accompanying documentation in the MDL website for this dataset: questionnaire.

- Name of the file in the MDL: **f1_firm_baseline_raw_data.dta**

Midline firm survey dataset

- Accompanying documentation in the MDL website for this dataset: questionnaire.

- Name of the file in the MDL: **f2_firm_midline_raw_data.dta**

---

[3]Dataset available in MDL (Global Indicators Department, Enterprise Analysis Unit, 2017).
[4]Datasets available in MDL (Crépon and Premand, 2024a).

Apprentice midline dataset

- Accompanying documentation in the MDL website for this dataset: questionnaire.

- Name of the file in the MDL: *f3_apprentice_midline_raw_data.dta*

Dataset on the trades of firms

- No documentation needed for this dataset, as it is self-explanatory in its content.

- Name of the file in the MDL: *f4_trade_of_firms.dta*

Youth surveys: baseline, midline, and endline datasets[5]

- Are these datasets included in the reproducibility package: no.

- Datasets are available in this MDL website

- Instructions on accessing these datasets in the MDL website: click on the **get microdata** tab. Register in the data portal and download the folder named **Data in Stata14 Format**.

- After downloading, please locate the six .dta datasets in the **Raw datasets/Youth** subfolder.

Baseline youth survey dataset

- Accompanying documentation in the MDL website for this dataset: questionnaire.

- Name of the file in the MDL: *y1_raw_youth_baseline.dta*

Midline youth survey dataset

- Accompanying documentation in the MDL website for this dataset: questionnaire.

- Name of the file in the MDL: *y2_raw_youth_midline.dta*

Endline youth survey dataset

- Accompanying documentation in the MDL website for this dataset: questionnaire.

---

[5]Datasets available in MDL (Crépon and Premand, 2024b).

- Name of the file in the MDL: ***y3_raw_youth_endline.dta***

Cote d'Ivoire national employment survey for year 2013

- Documentation available for this dataset: this website contains its respective questionnaire.

- Name of the file in the MDL: ***y4_employment_survey.dta***

Dataset on the youths that registered and started the apprenticeship

- No documentation needed for this dataset, as it only has 3 variables: ID of youths, number of youths that registered to the PEJEDEC apprenticeship program, and number of youths that actually started the program.

- Name of the file in the MDL: ***y5_youth_starting_apprship.dta***

Dataset on costs of the intervention per youth

- No documentation needed for this dataset, as it is self-explanatory in its content.

- Name of the file in the MDL: ***y6_program_costs.dta***

Firm listing for year 2021 [6]

- Are these datasets included in the reproducibility package: no.

- Datasets are available in this MDL website

- Instructions on accessing these datasets in the MDL website: click on the ***get microdata*** tab. Register in the data portal and download the folder named ***Data in Stata14 Format***.

- After downloading, please locate the two .dta datasets in the **Raw datasets/Listing** subfolder.

Dataset on the firm listing in year 2021

- Accompanying documentation in the MDL website for this dataset: questionnaire.

- Name of the file in the MDL: ***l1_listing_firm_data.dta***

---

[6]Datasets available in MDL (Crépon et al., 2024).

Dataset with the grid of the areas that were visited in the 2021 listing

- No documentation needed for this dataset, as it is self-explanatory in its content.

- Name of the file in the MDL: ***l2_grid_visited_areas.dta***

A summary on the instructions for accessing all datasets are in the data appendix of this document.

Contact

For any inquiries or assistance related to this reproducibility package, please contact Patrick Premand at ppremand@worldbank.org, Bruno Crépon at bruno.crepon@ensae.fr, and Daniel Corredor Vallejo at dcorredorvallejo@worldbank.org.

References

CRÉPON, B., E. LESTANT, AND P. PREMAND (2024): *Youth Employment and Skills Development Project - Apprenticeship Firm Listing 2021 [Data set]*, World Bank, Development Data Group.
https://doi.org/10.48529/KFCZ-6T55.

CRÉPON, B. AND P. PREMAND (2024a): *Youth Employment and Skills Development Project - Apprenticeship Firms Surveys 2014-2016 [Data set]*, World Bank, Development Data Group.
https://doi.org/10.48529/W2SZ-D257.

——— (2024b): *Youth Employment and Skills Development Project - Apprenticeship Youth Surveys 2014-2018 [Data set]*, World Bank, Development Data Group.
https://doi.org/10.48529/766E-B621.

GLOBAL INDICATORS DEPARTMENT, ENTERPRISE ANALYSIS UNIT (2017): *Enterprise Survey 2016 [Data set]*, World Bank, Development Data Group.
https://doi.org/10.48529/JA3S-0T37.

Appendix: Summary on instructions for accessing datasets online

Dataset	Name of dataset in .dta format	Dataset is in the package (yes or no)	Link to access the datasets and documentation online	Instructions to download the dataset	Where to save the data in the reproducibility package
Enterprise National Survey in Côte d'Ivoire for year 2016	***C-te-d-Ivoire-2016-full-data.dta***	No	Dataset at WB Enterprise survey data portal and documentation at World Bank Microdata Library	Register in the WB Enterprise Data portal. Then search and download a zip folder named *Côte d'Ivoire-2016-full data.*	Open the zip folder and save ***C-te-d-Ivoire-2016-full-data.dta*** to the **Raw datasets/Enterprise survey** subfolder
Baseline firm survey dataset	***f1_firm_baseline _raw_data.dta***	No	Dataset and documentation at World Bank Microdata Library	Click on the ***get microdata*** tab and download the folder named ***Data in Stata14 Format***	Locate the .dta dataset at the **Raw datasets/Firms** subfolder
Midline firm survey dataset	***f2_firm_midline _raw_data.dta***	No	Dataset and documentation at World Bank Microdata Library	Click on the ***get microdata*** tab and download the folder named ***Data in Stata14 Format***	Locate the .dta dataset at the **Raw datasets/Firms** subfolder
Apprentice midline dataset	***f3_apprentice_midline _raw_data.dta***	No	Dataset and documentation at World Bank Microdata Library	Click on the ***get microdata*** tab and download the folder named ***Data in Stata14 Format***	Locate the .dta dataset at the **Raw datasets/Firms** subfolder
Dataset on the trades of firms	***f4_trade_of_firms.dta***	No	Dataset at World Bank Microdata Library	Click on the ***get microdata*** tab and download the folder named ***Data in Stata14 Format***	Locate the .dta dataset at the **Raw datasets/Firms** subfolder

Dataset	Name of dataset in .dta format	Dataset is in the package (yes or no)	Link to access the datasets and documentation online	Instructions to download the dataset	Where to save the data in the reproducibility package
Baseline youth survey dataset	*y1_raw_youth _baseline.dta*	No	Dataset and documentation at World Bank Microdata Library	Click on the **get microdata** tab and download the folder named **Data in Stata14 Format**	Locate the .dta dataset at the **Raw datasets/Youth** subfolder
Midline youth survey dataset	*y2_raw_youth _midline.dta*	No	Dataset and documentation at World Bank Microdata Library	Click on the **get microdata** tab and download the folder named **Data in Stata14 Format**	Locate the .dta dataset at the **Raw datasets/Youth** subfolder
Endline youth survey dataset	*y3_raw_youth _endline.dta*	No	Dataset and documentation at World Bank Microdata Library	Click on the **get microdata** tab and download the folder named **Data in Stata14 Format**	Locate the .dta dataset at the **Raw datasets/Youth** subfolder
Côte d'Ivoire national employment survey for year 2013	*y4_employment _survey.dta*	No	Documentation at this website and dataset at World Bank Microdata Library	Click on the **get microdata** tab and download the folder named **Data in Stata14 Format**	Locate the .dta dataset at the **Raw datasets/Youth** subfolder
Dataset on the youths that registered and started the apprenticeship	*y5_youth_starting _apprship.dta*	No	Dataset at World Bank Microdata Library	Click on the **get microdata** tab and download the folder named **Data in Stata14 Format**	Locate the .dta dataset at the **Raw datasets/Youth** subfolder
Dataset on costs of the intervention per youth	*y6_program_costs.dta*	No	Dataset at World Bank Microdata Library	Click on the **get microdata** tab and download the folder named **Data in Stata14 Format**	Locate the .dta dataset at the **Raw datasets/Youth** subfolder

Dataset	Name of dataset in .dta format	Dataset is in the package (yes or no)	Link to access the datasets and documentation online	Instructions to download the dataset	Where to save the data in the reproducibility package
Dataset on the firm listing in year 2021	*l1_listing_firm_data.dta*	No	Dataset and documentation at World Bank Microdata Library	Click on the ***get microdata*** tab and download the folder named ***Data in Stata14 Format***	Locate the .dta dataset at the **Raw datasets/Listing** subfolder
Dataset with the grid of the areas that were visited in the 2021 listing	*l2_grid_visited_areas.dta*	No	Dataset at World Bank Microdata Library	Click on the ***get microdata*** tab and download the folder named ***Data in Stata14 Format***	Locate the .dta dataset at the **Raw datasets/Listing** subfolder