

From Training to Earning: The Medium-Term Impact of Dual Apprenticeships on Youth Employment

Abstract

1. Introduction to the reproducibility package

This repository contains the code to reproduce all the exhibits include in the "From Training to Earning: The Medium-Term Impact of Dual Apprenticeships on Youth Employment" paper and supplementary material (not included in the paper). The analysis data production and analysis code files are located in the main folder, while specific folders include the ado files needed to run the code, and the files needed to process each round of data.

Before running the code to produce the tables and figures for the paper, it is important to follow the next steps to finish the set up.

1. Make sure the raw datasets are stored properly in the correct directories (See 2. and 3. for more details).
2. Update lines 101 and 105 in the Main.do according to the location of the data folder and the code folder, respectively

2. Folder Structure

The master do-file will create the Data folder structure if not already present in the chosen directory stored in the user global. If that happens, the Main.do file will need a first run to create the folder structure. After that, the published datasets should be included in the corresponding folders (see 3.). As a third step, the Main.do file should be run a second time to produce the tables and figures.

2.1. Data

Data directory (chosen by the User)

```
|_ Data
  |_ SurveyData2016
    |_ 1_Raw
    |_ 3_Cleaned
  |_ SurveyData2018
    |_ 1_Raw
    |_ 3_Cleaned
  |_ SurveyData2019
    |_ 1_Raw
    |_ 3_Cleaned
  |_ SurveyData2021
    |_ 1_Raw
    |_ 3_Cleaned
```

2.2. Code and outputs

The code is structured as follows.

```
cdi-apprenticeship-long-term
|_ ado
|_ output
|   |_ tables
|   |_ figures
|_ SurveyData2016
|_ SurveyData2018
|_ SurveyData2019
|_ SurveyData2021
```

3. Data

The published version of the data used in the analysis is stored separately for each round. The next table includes each data file name, a link to download it, and where it should be located in the folder structure:

Dataset & Citation	Files
Youth Employment and Skills Development Project – Apprenticeship Youth Surveys 2014–2018 Crépon, B., & Premand, P. (2024). <i>Youth Employment and Skills Development Project - Apprenticeship Youth Surveys 2014–2018</i> [Data set]. World Bank, Development Data Group. https://doi.org/10.48529/766E-B621	SurveyData2016/1_Raw/y1_raw_youth_baseline.dta SurveyData2016/1_Raw/y2_raw_youth_midline.dta SurveyData2018/1_Raw/y3_raw_youth_endline.dta
Youth Employment and Skills Development Project – Apprenticeship Youth Surveys 2019 <i>(Forthcoming on World Bank Microdata Library)</i> Crépon, B., Lestant, E., & Premand, P. (2025). <i>Youth Employment and Skills Development Project - Apprenticeship Youth Surveys 2019</i> [Data set]. World Bank, Development Data Group.	SurveyData2019/1_Raw/y4_raw_youth_endline.dta
Youth Employment and Skills Development Project – Apprenticeship Youth Surveys 2021 <i>(Forthcoming on World Bank Microdata Library)</i> Crépon, B., Lestant, E., & Premand, P. (2025). <i>Youth Employment and Skills Development Project - Apprenticeship Youth Surveys 2021</i> [Data set]. World Bank, Development Data Group.	SurveyData2021/1_Raw/y5_raw_youth_endline.dta

Statement about Rights

- ☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ☒ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

4. Code structure

The following table details the code file names, a short description of their use, and their location with respect to the code folder structure (see 2.2.).

No.	File name	Description	location
1.	Main.do	Main code that runs all the processing and analysis. Runs 0_Master_2016_250608.do, 0_Master_2018_250608.do, 0_Master_2019_220507.do , and 0_Master_2021_220508.do	Main directory
2.	0_functions.do	Sets up the functions needed to output most tables	Main directory
3.	0_globals.do	Sets up all variable lists and column titles used in tables	Main directory
4.	1a_prep_wide_panel.do	Prepares the dataset used to get estimates for each round sample	Main directory
5.	1b_prep_long_panel.do	Prepares the dataset used to get the estimates for the pooled sample	Main directory
6.	2_tables_and_figures.do	Produces tables and figures included as exhibits in the paper and supplementary tables (not included in the paper)	Main directory
7.	0_Master_2016_250608.do	Main code for the 2014 and 2016 rounds. Merges rounds 2014 and 2016 and creates variables	SurveyData2016
8.	ano_youth_random_baseline_midline_Final.do	creates variables for rounds 2014 and 2016	SurveyData2016

No.	File name	Description	location
9.	recode youth midline selfemp var.do	Recodes self-employment for rounds 2014 and 2016	SurveyData2016
10.	0_Master_2018_250608.do	Main code for the 2018 round. Cleans and creates variables	SurveyData2018
11.	1_cleaning_by_module.do	Cleans and creates variables	SurveyData2018
12.	2_cleaning_recode_occupation.do	Cleans and creates variables related to occupation classification	SurveyData2018
13.	3_outcome_competence.do	Cleans and creates variables related to skills	SurveyData2018
14.	4_outcome_formation.do	Cleans and creates variables related to formation	SurveyData2018
15.	5_outcome_activity.do	Cleans and creates variables related to earnings	SurveyData2018
16.	0_Master_2019_220507.do	Main code for the 2019 round. Cleans and creates variables	SurveyData2019
17.	1_cleaning_by_module.do	Cleans and creates variables	SurveyData2019
18.	2_cleaning_recode_occupation.do	Cleans and creates variables related to occupation classification	SurveyData2019
19.	3_outcome_competence.do	Cleans and creates variables related to skills	SurveyData2019
20.	4_outcome_formation.do	Cleans and creates variables related to formation	SurveyData2019
21.	5_outcome_activity.do	Cleans and creates variables related to earnings	SurveyData2019
22.	0_Master_2021_220508.do	Main code for the 2021 round. Cleans and creates variables	SurveyData2021
23.	1_cleaning_by_module.do	Cleans and creates variables	SurveyData2021
24.	2_cleaning_recode_occupation.do	Cleans and creates variables related to occupation classification	SurveyData2021
25.	3_outcome_competence.do	Cleans and creates variables related to skills	SurveyData2021
26.	4_outcome_formation.do	Cleans and creates variables related to formation	SurveyData2021

No.	File name	Description	location
27.	5_outcome_activity.do	Cleans and creates variables related to earnings	SurveyData2021

5. List of exhibits

All figures and tables are produced in 2_tables_and_figures.do. Tables 1 to A6 are detailed here and included in the manuscript. Tables S1 - S18 are part of the supplementary material that is not included in the manuscript.

Output type and number	Output file name	Codeline
Figure 1	Figure1_Cross_time_tot_income_per_treat.png	line 74
Table 1	Table1_multiFGT.tex	line 81
Table 2	Table2_skill.tex	line 87
Table 3	Table3_income.tex	line 93
Table 4	Table4_employ.tex	line 99
Figure A1	FigureA1_Cross_time_pov_spell_treat_effect_fig2a.png	line 143
Figure A2	FigureA2_Distribution_Income_E2.png FigureA2_Distribution_Income_E3.png FigureA2_Distribution_Income_E4.png FigureA2_Distribution_Income_pooled.png	line 208 & 273
Figure A3	FigureA3_bcratio_updated_v3.png	line 358
Table A1	TableA1_Balancing_youth.tex	line 422
Table A2	TableA2_Balancing_youth_attr.tex	line 632
Table A3	TableA3_MHT.tex	line 788
Table A4	TableA4_index_breakdown.tex	line 926
Table A5	TableA5_hours.tex	line 942
Table A6	TableA6_hourly.tex	line 947
Table A7	TableA7_sub_asp.tex	line 952
Table A7	TableA8_obj_asp.tex	line 957