

README for “Do Workfare Programs Live Up to Their Promises? Experimental Evidence from Côte d’Ivoire”

Overview

This replication package contains the data and code necessary to reproduce the results in the paper “Do Workfare Programs Live Up to Their Promises? Experimental Evidence from Côte d’Ivoire”. The code is written in Stata and R.

The process involves two main master files: 1. `Global_Master.do`: A Stata script that handles data preparation and generates the main tables and figures for the impact evaluation. 2. `MASTER_2_ML.R`: An R script that performs the machine learning analysis to assess impact heterogeneity and targeting improvements.

The replicator should expect the Stata code to run for approximately 11 hours and the R code for over 20 hours on a high-performance machine. Due to the computational intensity of the machine learning component, we recommend using a machine with multiple cores (the code is set to use 15 by default).

Data Availability and Provenance Statements

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.

Summary of Availability

- All data **are** publicly available, but are not included in the reproducibility package.

The data used in this study were collected via surveys. The table below summarizes the primary survey datasets.

Data Name	Data Files	Location	Access	URL
Raw baseline Survey Data	<code>raw_data_baseline.dta</code>	<code>data_cleaning/Raw data/Baseline/</code>	Publicly accessible, not included in this package	World Bank Microdata Library
Raw midline Survey Data	<code>raw_data_midline_h.dta</code> , <code>raw_data_midline_i.dta</code>	<code>data_cleaning/Raw data/Midline/</code>	Publicly accessible, not included in this package	World Bank Microdata Library
Raw endline	<code>raw_data_endline.dta</code>	<code>data_cleaning/</code>	Publicly	World Bank

Data Name	Data Files	Location	Access	URL
Survey Data	ne.dta	Raw data/Endline/	accessible, not included in this package	Microdata Library

The URLs in the table above include the respective questionnaires. Baseline and Midline surveys were paper based. Endline survey was conducted with CAPI (Computer-Assisted Personal Interviewing), using Survey Solutions platform hosted by the World Bank. While no script is available for this, the instrument is provided as designed on the Designer platform (download from platform) and reflects all programming aspects such as filters and notes to enumerators. A paper version of the endline instrument is also provided.

Cleaned datasets

The raw baseline, midline, and endline survey data are processed in the data cleaning folder, using the cleaning do-file data_cleaning/Cleaning code/Master cleaning.do, and deposited on the data/deidentified/ folder. In order to simplify the computational process, the master do-file that reproduces the outputs of this paper starts after this cleaning step, with the datasets below, which are also available in the World Bank Microdata Library.

Data Name	Data Files	Location	Access	URL
Baseline Survey Data	Baseline_i_hh_clean_var.dta, ENSETE_2013_17 Septembre2014_nonmissing.dta	data/deidentified/	Publicly accessible, not included in this package	World Bank Microdata Library
Midline Survey Data	Midline_hh_clean.dta, Midline_i_clean_var_est.dta	data/deidentified/	Publicly accessible, not included in this package	World Bank Microdata Library
Endline Survey Data	Endline_i_hh_clean_est.dta	data/deidentified/	Publicly accessible, not included in this package	World Bank Microdata Library

Note on availability of intermediate ML data produced in R

Given the high computational demands for reproducing the R code (computational requirements and recommendation to use a computer with multiple cores shown below), the authors included the intermediate data from the Machine Learning (ML) process (in the folder ML_reproduce_rep1) directly in this reproducibility package. With this, the replicator can avoid running the R code and start directly with the cleaning and estimations steps in Stata to obtain the tables/figures of the paper.

Computational requirements

Software Requirements

- The replication package contains two master programs to run all analyses.
- **Stata version 15**
- **R version 3.6.4+**
 - The R code relies on packages and functions from the following sources, which are handled within the scripts (latest version in this reproducibility package):
 - MLInference
 - Bob_Emploi_RCT
 - Helper scripts `util.R`, `helper_functions.R`, and `methods_functions.R` manage packages and define necessary functions.

Controlled Randomnes

- No Pseudo random generator is used in the analysis described here.

Memory, Runtime, Storage Requirements

Summary

Approximate time needed to reproduce the analyses on a high-performance machine (e.g., Apple M2 Max with 96GB RAM):

- 1-3 days

Details: * **Stata analysis (Global_Master.do):** ~11 hours. * **R analysis (MASTER_2_ML.R):** 20+ hours. * **WARNING:** The ML code is computationally intensive. We strongly recommend using a computer with multiple cores (default is set to 15). As mentioned above, intermediate data from the ML process is included in the reproducibility package to allow for the reproduction of the final and tables without running the ML R script.

Approximate storage space needed:

- 250 MB - 2 GB

Details

The code was last run on an **Apple M2 Max with 96GB of RAM**. The R code is configured to take advantage of multi-core processing to reduce runtime.

Description of programs/code

The code is organized into two main workflows, controlled by master files.

- **Global_Master.do**: This is the main Stata master file. It calls other Stata scripts to perform data preparation and generate all tables and figures related to the primary impact evaluation.
 - MASTER_1_DATASETS.do: Prepares and merges datasets.
 - MASTER_3_TABLES_AND_FIGURES.do: Generates final exhibits.
- **MASTER_2_ML.R**: This is the main R master file for the machine learning analysis. It should be run after the initial data preparation in Stata. It produces the heterogeneity analysis, predicted impacts, and targeting simulations.

License for Code

The code is licensed under a Modified BSD3 license. See LICENSE.txt for details.

Instructions to Replicators

To reproduce all results, follow these steps in sequence:

0. Placing the data in the respective folders.

- Download data from the sources mentioned in this document and place them in the indicated locations.

1. Reproducing the data cleaning package for baseline, midline, and endline raw datasets.

- Open data_cleaning/Cleaning code/Master_cleaning.do
- Update the cleaning directory path in line 12.
- Run the do-file.

2. Set up Directory Paths

- In Global_Master.do (Stata), update the global macro for your main directory path.
- In MASTER_2_ML.R (R), update the directory paths on lines 53 and 57 to point to your working directory and the data folder.

3. Run Stata Data Preparation

- Open Global_Master.do in Stata.
- Run the first part of the script which calls MASTER_1_DATASETS.do. This will create the intermediate datasets that are required for the R script.

4. Run R Machine Learning Analysis

This step uses the MASTER_2_ML.R script.

WARNING: The code is quite demanding in computer power and will probably take weeks to run on a normal PC. We strongly recommend using computers with multiple cores (default set for 15) to run the proxy predictions. As mentioned above, intermediate data

from the ML process is included in the reproducibility package to allow for the reproduction of the final and tables without running this step.

ML Folder Structure

Ensure your working directory has the following structure. The scripts will create the results directory and its sub-folders if they don't exist.

```
Package/
├── codes/
│   ├── Auxiliary/
│   │   ├── util.R
│   │   ├── helper_functions.R
│   │   └── methods_functions.R
│   └── Main/
│       └── MASTER_2_ML.R
├── data/
└── results/
    ├── experiment_name/
    │   ├── outcome/
    │   ├── winsor_num/
    │   └── treatment_arm/
    ├── B_and_S/
    ├── settings/
    ├── stata/
    └── samples/
```

ML Input Data

The following datasets, generated in Step 3, are required in your data folder: *

Baseline_Mid_End_200522.dta * other_outcome_data.dta

Reproducing ML Results

1. Open MASTER_2_ML.R file located in the codes/Main folder.
2. Set the workflow you will use with Dropbox on line 51 (Through the API (set dropbox_API=TRUE) or local folder (set dropbox_API=FALSE)).
3. Confirm the working directory and data/output directory paths are set correctly (lines 53 and 57).
4. Run the entire script. This will generate the necessary .dta files with ML predictions in the data/ML_reproduce folder.
5. **Specify the set of parameters:** Upon running the file, you need to enter a set of estimation parameters from the 18 settings outlined in the table below to execute via a command line prompt. You need to either specify one set of parameters at a time, or to reproduce all results, you must enter 0.

ML Prediction Settings

For example, choosing chosen_setting = 1 will produce predictions for y0_m (log monthly earnings at midline) under the specification detailed in the first row.

Setting #	Target output	treatment arm	folder name	winsorization quantile	Table in paper that uses the predictions	Set of covariates
1.	y0_m	c(1,2,3)	ALL	97	Table 6, 7, 8, 9, 10, A7; Figure 2, 3, B4, B5	SMALL
2.	y0_e	c(1,2,3)	ALL	97	Table 6, 10, A7; Figure 2, 3, B4	SMALL
3.	y1_m	c(1,2,3)	ALL	97	Table 6, 10, A7, A12; Figure 4, B4, B5, B6	SMALL
4.	y1_e	c(1,2,3)	ALL	97	Table 6, 10, A7; Figure B4, B6	SMALL
5.	s0_m	c(1,2,3)	ALL	97	Table A13, Table A14	SMALL
6.	s0_e	c(1,2,3)	ALL	97	Table A13	SMALL
7.	s1_m	c(1,2,3)	ALL	97	Table A13	SMALL
8.	s1_e	c(1,2,3)	ALL	97	Table A13	SMALL
9.	y0_e	c(1)	WET	97	Table A9	SMALL
10.	y0_e	c(2)	SET	97	Table A9	SMALL
11.	y0_e	c(3)	PW	97	Table A9	SMALL
12.	y1_e	c(1)	WET	97	Table A10	SMALL
13.	y1_e	c(2)	SET	97	Table A10	SMALL
14.	y1_e	c(3)	PW	97	Table A10	SMALL
15.	y0_m	c(1,2,3)	ALL	97	Table A8	LARGE

Setting #	Target output	treatment arm	folder name	winsorization quantile	Table in paper that uses the predictions	Set of covariates
16.	y0_e	c(1,2,3)	ALL	97	Table A8	LARGE
17.	y1_m	c(1,2,3)	ALL	97	Table A8	LARGE
18.	y1_e	c(1,2,3)	ALL	97	Table A8	LARGE

Optional ML Changes

- You can change the number of cores used for proxy predictions (line 474) and for Chernozhukov's estimators (line 567) to match your machine's capabilities.

5. Generate Final Tables and Figures

- Once the R script has been run for all necessary settings and the prediction files are generated, run the final part of `Global_Master.do` (which calls `MASTER_3_TABLES_AND_FIGURES.do`). This script uses the outputs from both the Stata and R analyses to generate all the final tables and figures in the Output folder.

List of tables and programs

- All tables and figures in the paper are reproduced by the code.

Figure/Table #	Program	Output file	Note
Table 1: Baseline Summary Statistics and Balance Checks	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_1_balance_weighted_prueba	
Table 2: Characteristics of Applicants and National Population of Urban Youths	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_2_ENSETE_comp	
Table 3: Impacts during and post program, employment and hours	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_3_main_itt	
Table 4: Impacts during and post program, employment and hours	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_3b	
Table 5: Impacts during and post	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_5_main_itt_behavior_wel	

Figure/Table #	Program	Output file	Note
program, well-being and behaviors		lbeing	
Table 6: Impacts post program on skills, investments in self-employed activities and search for wage jobs	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A5_it_t_bhv_skills	
Table 7: Heterogeneity in impacts on earnings during and post program, machine learning results	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_6_ml_v3	Requires ML script output
Table 8: Baseline characteristics of the bottom and top quartiles of the distribution of predicted impacts on (ln) earnings during program	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_7_ml_charac_v3	Requires ML script output
Table 9: Impacts during and post program on main outcomes, by quartile of predicted impacts on (ln) earnings during the program	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_8_during_post_v3	Requires ML script output
Table 10: Impacts post program on intermediate outcomes, by quartile of predicted impacts on (ln) earnings during the program	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_9_ml_post_bis_v3	Requires ML script output
Table 11: Impacts on (ln) earnings and cost-benefit ratios	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_10_target_MLv3	Requires ML script output

Figure/Table #	Program	Output file	Note
under alternative targeting approaches			
Figure 1: Quantile treatment effects for (ln) earnings during and post program	MASTER_3_TABLES_AND_FIGURES.do	Figures/figure_1_quantile_all.png	
Figure 2: Quantile treatment effects for (ln) earnings post program, by treatment arm	MASTER_3_TABLES_AND_FIGURES.do	Figures/figure_B3_quantile_Tarm.png	
Figure 3: Group average treatment effects (GATES) for (ln) earnings	MASTER_3_TABLES_AND_FIGURES.do	Figures/figure_2_group_ates_MLv3.png	Requires ML script output
Figure 4: Predicted impact on (ln) earnings during vs post program	MASTER_3_TABLES_AND_FIGURES.do	Figures/figure_3_scatter_cate_v3.png	Requires ML script output
Figure 5: CATE at Midline by predicted BL earnings	MASTER_3_TABLES_AND_FIGURES.do	Figures/graph_predbase_vs_predcate_mid.png	
Table A1: Overview of Public Works Programs	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A1_programs	
Table A2: Impacts during and post program, robustness to alternative definition of earnings	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_4_robustness_it	
Table A3: Estimated impacts during and post program on economic outcomes, with baseline controls	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A2_controls	
Table A4: Estimated impacts during and post program on	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A3	

Figure/Table #	Program	Output file	Note
economic outcomes, without clustering standard errors			
Table A5: Impacts during and post program, full sample at follow-up	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A4	
Table A6: Impacts during and post program, LATE accounting for compliance in control between midline and endline	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A5	
Table A7: Estimated impacts during program on economic outcomes for household members other than the beneficiary	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_hh_itt	
Table A8: Estimated impacts during and post program on well-being index components	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A2_itt_wb_comp	
Table A9: Estimated impacts during and post program on behavior index components	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A3_itt_bhv_comp	
Table A10: Estimated impacts during and post program on risky behaviors	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A4_list_linear	
Table A11: Impacts during and post program on time use	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_app2b	

Figure/Table #	Program	Output file	Note
Table A12: Estimated impacts post program on aspirations and reservation wage	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_aspiration_itt	
Table A13: Comparison of Machine Learning algorithms to predict impacts on earnings during and post program	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A6_ML_methods_v3	Requires ML script output
Table A14: Heterogeneity in impacts on earnings during and post program, machine learning results for an extended set of covariates	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A7_ml_extcov_v3	Requires ML script output
Table A15: Estimated impacts on (ln) earnings post program, by treatment arms	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A8_ml_tarms_v3	Requires ML script output
Table A16: Estimated impacts on earnings (in levels) post program, by treatment arms	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A9_ml_tarms_lv1_v3	Requires ML script output
Table A17: Baseline characteristics of the bottom and top quartiles of predicted impacts on earnings (in levels) during program	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A10_ml_charac_lv1_v3	Requires ML script output
Table A18: Estimated impacts during and post	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A11_ml_svg_v3	Requires ML script output

Figure/Table #	Program	Output file	Note
program on savings			
Table A19: Baseline characteristics of the bottom and top quartiles of predicted impacts on (ln) savings during program	MASTER_3_TABLES_AND_FIGURES.do	Tables/Table_A12_m1_charac_svg_v3	Requires ML script output
Figure B1: Cumulative distribution of earnings during and post program	MASTER_3_TABLES_AND_FIGURES.do	Figures/graph_cum_dist_income.png	
Figure B2: Reported baseline earnings vs Predicted impacts during and post program	MASTER_3_TABLES_AND_FIGURES.do	Figures/graph_base_vs_pred.png	
Figure B3: Cost-effectiveness ratios over time under alternative targeting rules, depending on the sustainability of post-program impacts	MASTER_3_TABLES_AND_FIGURES.do	Figures/Figure_B5_CE_v3.png	Requires ML script output
Table F1: Baseline variables used in Machine Learning algorithms	n/a	(Directly in LaTeX code)	Not produced by scripts
Figure F1: Relation between predictions and actual earnings (in level) (Random forest)	MASTER_3_TABLES_AND_FIGURES.do	Figures/figure_B6_act_vs_pred_rf_lv1.pdf	Requires ML script output
Table G1: Summary of weights used with midline data	n/a	(Directly in LaTeX code)	Not produced by scripts
Table G2: Summary of weights used with	n/a	(Directly in LaTeX code)	Not produced by scripts

Figure/Table #	Program	Output file	Note
endline data			

Reproducibility when using ML data

The estimations for some tables/figures that require machine learning data can be run in two ways: - 1 - Starting from scratch, generating the ML intermediate data using the machine learning algorithms in R, or - 2 - Using the intermediate ML data provided directly by the authors to produce the final exhibits.

Some differences in estimations between running option 1 and option 2 can arise due to differences in operating systems (e.g., Mac vs. Windows) or package versions. However, such differences are minor and do not affect the interpretation of the results. The main tables/figures where there are minor differences between results from option 1 and option 2 are the following:

- Table 7
- Table 8
- Table 9
- Table 10
- Table 11
- Figure 3
- Figure 4
- Figure 5

For full transparency, screenshots documenting these minor differences are included in the *Reproducibility Verification Report* accompanying this package. The report can be found in [this link](#).

References

All data included in this package is publicly available in the following links.

- Projet Emploi Jeune et Développement des Compétences (PEJEDEC) Impact Evaluation - Public Works Baseline Survey, 2013. Citation: Marianne Bertrand, Bruno Crépon, Alicia Marguerie, Patrick Premand. Côte d'Ivoire - Projet Emploi Jeune et Développement des Compétences (PEJEDEC) Public Works Impact Evaluation - Baseline Survey, 2013 (PEJEDEC PWP BL 2013). Ref: CIV_2013_PEJEDEC-PWP-BL_v01_M. Downloaded from <https://microdata.worldbank.org/index.php/catalog/6774>
- Projet Emploi Jeune et Développement des Compétences (PEJEDEC) Impact Evaluation - Public Works Midline Survey, 2013. Citations; Marianne Bertrand, Bruno Crépon, Alicia Marguerie, Patrick Premand. Côte d'Ivoire - Projet Emploi Jeune et Développement des Compétences (PEJEDEC) Public Works Impact

Evaluation - Midline Survey, 2013 (PEJEDEC PWP ML 2013). Ref:
CIV_2013_PEJEDEC-PWP-ML_v01_M. Downloaded from
<https://microdata.worldbank.org/index.php/catalog/6775>

- Projet Emploi Jeune et Développement des Compétences (PEJEDEC) Public Works Impact Evaluation - Endline Survey, 2015. Citation: Source: Marianne Bertrand, Bruno Crépon, Alicia Marguerie, Patrick Premand. Côte d'Ivoire - Projet Emploi Jeune et Développement des Compétences (PEJEDEC) Public Works Impact Evaluation - Endline Survey, 2015 (PEJEDEC PWP EL 2015). Ref:
CIV_2015_PEJEDEC-PWP-EL_v01_M. Downloaded from
<https://microdata.worldbank.org/index.php/catalog/6776>