

READ ME

Overview

This file describes the procedures needed to replicate the results for the paper “Turning Inward and/or Outward: Which Socioemotional Skills Pay for Agribusiness Entrepreneurs in Nigeria?” by Smita Das, Clara Delavallade, Ayodele Fashogbon, and Sreelakshmi Papineni. The code in this replication package constructs the analysis files using Stata.

Data Availability Statements

Statement about Rights

The data from an impact evaluation contained in this package was collected by the authors and the authors have legitimate access to and permission to use the data in this manuscript.

Summary of Availability

The authors have permission to redistribute/ publish the data contained within this replication package. All data has been de-identified of personal information as per IRB requirements. The data will be embargoed until publication of the paper in a journal.

Data Sources

Baseline and Screening Survey

- **Filename:** APPEALS_EndlineFieldList_08112023.dta
- **Source:** Screening data collected by relevant authorities as part of program screening activities; baseline data collected by the authors through original field surveys
- **URL:** Not applicable
- **Access year:** 2023

Endline Survey

- **Filename:** APPEALS_endline.dta
- **Source:** Original field surveys conducted by the authors

- **URL:** Not applicable
- **Access year:** 2024

Note: Screening data were collected between May–December 2019, baseline data between June–October 2020, endline data between September–October 2023, with mop-up data collected between April–July 2024.

Replication folder directory

This project is organized into several key folders. “data_raw” holds the unprocessed, original datasets. The “data_intermediate” folder stores the cleaned and pre-processed data, and “data_final” contains the clean datasets ready for analysis. “output” includes all the results for tables generated (in both the main body of the text and the appendix). The “code” folder contains the do-files used for data processing, analysis, and generating output.

The replicator can run all cleaning and analysis files from start to end without interruption using the master do-file `code/master.do`. The replicator will need to first define a personal directory path in `code/master.do` before running the program.

Do-files

master.do runs the code to generate all relevant files for analysis, and runs each of the do-files for data importing, cleaning, and analysis (further described below).

00_setup.do runs the code to set the directory.

10_import_clean.do runs the code to construct the data used in the analysis. This includes cleaning the data and merging survey rounds (screening, baseline, and endline).

20_construct_vars.do runs the code to create outcome variables used in the analysis.

30_analysis_main.do runs the code to generate the main tables (Tables 2-10) in the paper.

40_analysis_appendix.do runs the code to generate the tables (Tables C1-C12) presented in the supplementary appendix.

50_analysis_baltab.do runs the code to generate Table 1 presented in the main text and Table A.2 in the supplementary appendix.

Data

All results reported in the paper can be reproduced by running the Stata do-files contained in the `code` folder of the replication package. The file `30_analysis_main.do` generates the results presented in the main text (Tables 2–10), while `40_analysis_appendix.do` produces the appendix tables (Tables C.1–C.12). The file `50_analysis_baltab.do` generates Table 1 and Table A.2. To replicate the results, these do-files draw on multiple datasets and should run from start to end without interruption. To perform the analysis, we rely on the following datasets:

1. **APPEALS_EndlineFieldList_08112023.dta** (includes data from the screening and baseline rounds, with variables used for ANCOVA analysis, norm analysis, and other baseline controls).
2. **WYEP_endline.dta** (individual-level data from the endline survey).

All results (main and appendix) will be stored in the folder “output”. Main tables are stored in the “tables_main” folder, appendix tables are stored in the “tables_appendix” folder, and figures from the residual–residual analysis are stored in the “figures” folder within the output directory.

Table construction and presentation

Tables were generated from Stata output and exported as `.xlsx/.csv` files, which are included in the replication package. For manuscript preparation, these files were formatted and converted into LaTeX tabular code using a standard spreadsheet-to-LaTeX table conversion tool (TableConvert), with additional manual formatting using ChatGPT to accommodate table size, layout, and notes. No numerical values were altered during this process.

List of Exhibits

The provided code reproduces:

- ☐ All numbers provided in text in the paper
- ☒ All tables and figures in the paper
- ☐ Selected tables and figures in the paper only

Exhibit Identification

Exhibit name	Output filename	Script	Notes
Table 1	table1_nrmd.xlsx; table1_ttest.xlsx	50_analysis_baltab.do	Raw outputs stored in output/tables_main. Formatted table stored in output/tables_formatted.
Table 2	table2.csv	30_analysis_main.do	Stored in output/tables_main.
Table 3	table3.csv	30_analysis_main.do	Stored in output/tables_main.
Table 4	table4.csv	30_analysis_main.do	Stored in output/tables_main.
Table 5	table5.csv	30_analysis_main.do	Stored in output/tables_main.
Table 6	table6.csv	30_analysis_main.do	Stored in output/tables_main.
Table 7	table7.csv	30_analysis_main.do	Stored in output/tables_main.
Table 8	table8.csv	30_analysis_main.do	Stored in output/tables_main.
Table 9	table9.csv	30_analysis_main.do	Stored in output/tables_main.
Table 10	table10.csv	30_analysis_main.do	Stored in output/tables_main.
Table A.1	tableA1.pdf	Not Applicable	Stored in output/tables_appendix. It has been constructed using administrative records provided by the implementing partner, and beneficiary survey data collected during the training sessions.
Table A.2	tableA2_nrmd.xlsx; tableA2_ttest.xlsx	50_analysis_baltab.do	Raw outputs stored in output/tables_appendix. Formatted table stored in output/tables_formatted.

Exhibit name	Output filename	Script	Notes
Table A.3	tableA3.pdf	Not Applicable	Stored in output/tables_appendix. It draws on multiple external data sources. Eligibility statistics are obtained from the National Bureau of Statistics (NBS) National Living Standards Survey (NLSS) 2020. Census figures are based on the NBS Projected Census 2019. Number of beneficiaries is drawn from administrative records of the APPEALS project.
Table C.1	tableC1a.csv; tableC1b.csv	40_analysis_appendix.do	Raw outputs stored in output/tables_appendix. The two output files are combined and formatted to generate the final table reported in the manuscript.
Table C.2	tableC2a.csv; tableC2b.csv	40_analysis_appendix.do	Raw outputs stored in output/tables_appendix. The two output files are combined and formatted to generate the final table reported in the manuscript.
Table C.3	tableC3.csv	40_analysis_appendix.do	Stored in output/tables_appendix.
Table C.4	tableC4.csv	40_analysis_appendix.do	Stored in output/tables_appendix.
Table C.5	tableC5.csv	40_analysis_appendix.do	Stored in output/tables_appendix.
Table C.6	tableC6.csv	40_analysis_appendix.do	Stored in output/tables_appendix.
Table C.7	tableC7.csv	40_analysis_appendix.do	Stored in output/tables_appendix.

Exhibit name	Output filename	Script	Notes
Table C.8	tableC8.csv	40_analysis_appendix.do	Stored in output/tables_appendix.
Table C.9	tableC9.csv	40_analysis_appendix.do	Stored in output/tables_appendix.
Table C.10	tableC10.csv	40_analysis_appendix.do	Stored in output/tables_appendix.
Table C.11	tableC11.csv	40_analysis_appendix.do	Stored in output/tables_appendix.
Table C.12	tableC12.csv	40_analysis_appendix.do	Stored in output/tables_appendix.
Figure 5.1–5.4	binscatter_ <outcome>.png (11 files)	40_analysis_appendix.do	Residual–residual binscatter figures. One PNG file is generated per outcome variable and saved in output/figures. The set of files together corresponds to Appendix Figures 5.1–5.4 in the manuscript.

Note: File names may not correspond exactly to exhibit numbers in the manuscript. The table above provides a one-to-one mapping between manuscript exhibits, output files, and the scripts used to generate them.

Computational requirements

Software Requirements

Stata (code was last run with version 17)

- ietoolkit
- ritest
- unique
- renvars
- reghdfe
- binscatter

- estout
- egenmore
- ietoolkit
- outreg2
- winsor

The program `code/master.do` will install all dependencies locally if not already installed.

Memory, Runtime, Storage Requirements

Approximate time needed to reproduce the analyses on a standard (2025) laptop machine: **20 minutes**. The replicator should expect the code to reproduce the main tables to run in approximately 10 minutes, and the code for the tables in the appendix in approximately 10 minutes.