

Do Informal Businesses with More Educated Owners Adopt Better Business Practices? Evidence from Central African Republic

By: Mohammad Amin*, Asif M. Islam**, and Debasmita Padhi***

This draft: November 4, 2025

Results replication

Step 1: Create a folder for storing all the data and results.

Step 2: Go to Enterprise Surveys website – www.enterprisesurveys.org. Click on the “Download Datasets” tab that you will see at the upper-right side of the screen. Next, click on “WB STAFF LOGIN” or else enter your login details. For the latter you need to be a registered user. Please register if necessary. Please contact Rose Gachina at DECEA if you need help with downloading the data. Once you are logged in, click on “SURVEY TYPE” on left side of the screen and select “Informal Enterprise Survey” from the drop-down menu. Next click on the “ECONOMY” tab just below and select Central African Republic from the drop-down menu. You should see “Central-African-Republic-2023-ISES-full-data.dta” link on your screen. Click on the link and then on “DOWNLOAD” link that you will see. Now complete the download and save the data in the folder created in Step 1 above. The data should be saved as “Central-African-Republic-2023-ISES-Full-data.dta” (this is the file name by default). Please note that this data cannot be stored in the Reproducibility Repository.

The above data was downloaded by the authors on 6/3/2024.

Step 3: Now, please go to line 4 in the do file provided. Replace “XX” with the full path (directory) of the folder created in Step 1. Now save the do file.

Step 4: Install fsum, cmp, ebalance, and teffects programs. That is, type the following commands in the Stata command window one at a time and then press Enter:

```
ssc install fsum
```

```
ssc install cmp
```

```
ssc install ebalance
```

Some of these programs may be already installed on your computer.

Another program you will need is “teffects”. This is a built-in program in Stata 17 (and Stata 16 and possibly earlier Stata versions). So, you don’t need to download this. You can check for this program by typing “which teffects” in the Stata command window and then press Enter.

The versions I have used are as follows:

Stata version used is Stata 17. The cmp version used is “cmp 8.7.7 31 August 2023”; ebalance version used is “version 1.5.4 Jens Hainmueller, Yiqing Xu 01/29/15”; teffects version used is “version 1.1.0 08apr2021”; R-Studio version used is “4.4.2 (2024-10-31 ucrt)”.

Step 5: Now run the do file from line 1 to line 429. The data is now ready to replicate the results.

Step 6: For replicating the graphs/figures, run the commands in the do file for the relevant figure. The commands for each figure are clearly marked and separated in the do file. It is best to run the codes figure by figure. Instructions are provided in the do file at the beginning of a graph or at the end of a command line (in green). Please follow these instructions.

Data for Figure 5 is taken from World Development Indicators (WDI), World Bank. It was accessed through DDP. Go to the WDI data. For country, select “Central African Republic”. For time period, please select 1990 to 2020. For the series, please select “Prevalence of HIV, total (% of population ages 15-49)” that has the following series code: SH.DYN.AIDS.ZS. Save the data in an excel and plot the bar graph.

Step 7: Figure 10 plots the coefficients shown in columns 1 and 3 in Table A10 in the Appendix. We have copy-pasted these coefficients in the excel file that is included with this package as “Figure10_Balance-plot-Data-for-R_v2.xlsx”.

To replicate Figure 10, you need to use R-Studio. Please run the R-script that is included with this package. The script name is “R-for-Fig10-Balance-plot_7-31-2025.R”. Open the script. In line 7, replace “XX” with the full path of the folder created in Step 1. Please note that unlike Stata, R uses forward slashes (/) in the directory/path name. Save the script. Now run the full script. Click on “Plots” on the right side of your screen. Increase the size of this window if you do not see the full graph.

Step 8: Replicating the tables in the manuscript:

- (i) Stata codes for all the tables are clearly marked and separated.
- (ii) For some tables, the output is saved in an excel file called “xml_7-21-2025.xls” in the folder created in Step 1 above. Open this excel file to see these tables. For other tables, the results should be read from the computer screen and matched to the ones in the table. The do file clearly specifies which tables are to be read from the screen. Please see the instructions at the beginning of each table or at the end of a command line (in green).
- (iii) Most tables have two Panels – A and B. These were manually combined in the manuscript. The tables saved in the “xml_7-21-2025.xls” file contain the Panel A and Panel B results for a table in two separate worksheets. For example, results in Panel A in Table 3 will be saved in the worksheet called “t3_A”, and the same for Panel B will be saved in “t3_B”. Likewise for other tables. It is best to generate the tables one by one so the instructions provided in the do file can be followed.
- (iv) The tables in the manuscript have been manually formatted. So, the results that you will see in the excel file will appear slightly different in terms of formatting, variable names, and column headings. But the results are exactly the same.
- (iv) Sometimes, the “cmp(.)” does not run first time. If this happens and an error shows, please run it a second time. It should work.

Third party candidates are allowed to run this reproducibility package.

I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Please feel free to email me at mamin@worldbank.org if you have any questions.
Thank you!