



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

First Submission: RR_CAF_2025_487

Camila Ayala

reproducibility@worldbank.org

December 4, 2025

This review verifies the reproducibility of the exhibits included in the paper “*Do Informal Businesses with More Educated Owners Adopt Better Business Practices? Evidence from Central African Republic*”.

Contents in this review:

1. Main findings
2. List of exhibits and reproducibility status
3. Reproduction Environment

Main findings

- **Every exhibit in the main section of the paper and 10 randomly selected exhibits in the appendix were reproduced accurately**
- The code was successfully executed on a new computer after:
 1. Downloading the dataset following the instructions provided in the README file.
 2. Changing the path in the do file *P99_CAF-Edu-MgmtPractice_11-4-2025.do* and running the code.
 3. Changing the path in the R script *R-for-Fig10-Balance-plot_7-31-2025.R* and running the code.
- Output stability was not verified due to the format of the outputs (copied from the Stata results window or stored in an .xls file format).
- The code takes approximately 6 minutes to run.
- We conducted our reproducibility analysis based on the paper shared by the authors by email on November 15th.
- **Verification Process and Data Handling:**
 - The reproducibility package relies on two types of data: open, and limited-access (available from the source website).
 - Open data is included in the public reproducibility package.
 - Reviewers used limited-access data downloaded directly from the Enterprise Surveys website to conduct the reproducibility verification. This data is not included in the public reproducibility package; however, detailed instructions on how to access them are provided in the README file.

- *data_hash_report.csv* lists the SHA256 hashes of all files in the Data folder to support data integrity checks. Users who acquire the limited-access data can use this file to verify that the data has not been altered.
- **Reproducibility Summary:**
 - **Data:** Open data are included in the package, while restricted data are not. For further details, please refer to the README file.
 - **Code:** All code files (from cleaning to analysis) are included in the reproducibility package.
 - **Outputs:** All outputs are generated by code included in the reproducibility package.
 - **Dependencies environment:** The reviewers created a new environment using the latest versions of dependencies available at the moment of the review.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Figure 1 Reproduced.** Verified against the mean values displayed in the Stata results window.
- **Figure 2 Reproduced.** Verified against the mean values displayed in the Stata results window.
- **Figure 3 Reproduced.** Verified against the percent values displayed in the Stata results window.
- **Figure 4 Reproduced.** Verified against the percent values displayed in the Stata results window.
- **Figure 5 Reproduced.** Verified against the file Figure5.xlsx.
- **Figure 6 Reproduced.** Verified against the mean values displayed in the Stata results window.
- **Figure 7 Reproduced.** Verified against the mean values displayed in the Stata results window.
- **Figure 8 Reproduced.** Verified against the mean values displayed in the Stata results window.
- **Figure 9 Reproduced.** Verified against the graph generated in Stata. Although there are some formatting differences, the densities in both figures are identical; therefore, the result is considered reproducible.
- **Figure 10 Reproduced.** Verified against the graph generated in R.
- **Figure 11 Reproduced.** Figures 11a to 11i were verified against the graphs generated in Stata. Although there are some formatting differences, all figures in the manuscript are identical to those generated by the code.
- **Table 1 Reproduced.** Verified against sheet t1 of xml_7-21-2025.xls. A minor labeling difference was identified: the manuscript reports “Profits (100,000 XFAs)”, while the table generated by the code reports “Profits (100,000 LCUs)”. This difference is present in all the tables. All numerical values are identical; therefore, the table is considered reproducible.

- **Table 2 Reproduced.** Verified against sheet t2 of `xml_7-21-2025.xls`. Some formatting and labeling differences were identified, but all numerical values are identical; therefore, the table is considered reproducible.
- **Table 3 Reproduced.** Verified against sheets t3_A and t3_B of `xml_7-21-2025.xls`. All numerical values are identical; therefore, the table is considered reproducible.
- **Table 4 Reproduced.** Verified against the estimates displayed in the Stata results window. Some values differ between the manuscript and the Stata output. However, these discrepancies occur only at the third decimal place, likely due to rounding; therefore, the table is considered reproducible.
- **Table 5 Reproduced.** Verified against sheet t5 of `xml_7-21-2025.xls`. Some formatting and labeling differences were identified, but all numerical values are identical; therefore, the table is considered reproducible.
- **Table 6 Reproduced.** Verified against the estimate values displayed in the Stata results window.
- **Table 7 Reproduced.** Verified against sheets t7_A and t7_B of `xml_7-21-2025.xls`. Minor formatting and labeling differences were identified, along with a change in the variable ordering; however, all numerical values are identical. Therefore, the table is considered reproducible.

Results in the Annex

For the Appendix, we did not review every exhibit. Instead, we randomly selected 10 exhibits from the remaining datasets to assess the appendix. Our review was based on those 10 exhibits. Since they were chosen randomly, we are operating under the assumption that if all randomly selected exhibits are reproducible, then the rest should be as well. The seed used to generate the random selection was 400379, the Stata version used was 18, and the exhibits selected were: 2, 3, 4, 5, 6, 7, 8, 9, 13, and 14.

- **Table A2 Reproduced.** Verified against the descriptive statistics displayed in the Stata results window.
- **Table A3 Reproduced.** Verified against the correlation matrix displayed in the Stata results window. The only difference is a minor rounding discrepancy in the correlation between Education and Started by owner (0.115 in the manuscript and 0.12 in Stata), therefore, the table is considered reproducible.
- **Table A4 Reproduced.** Verified against sheet tA4_1 of `xml_7-21-2025.xls`.
- **Table A5 Reproduced.** Verified against sheets A5_A and A5_B of `xml_7-21-2025.xls`. Some formatting and labeling differences were identified, but all numerical values are identical; therefore, the table is considered reproducible.
- **Table A6 Reproduced.** Verified against sheets A6_A and A6_B of `xml_7-21-2025.xls`. Minor formatting and labeling differences were identified. A slight discrepancy appears in the marginal effect of education in Panel B, column (4): the manuscript reports 0.079, while the Excel file reports 0.08. As this difference is minimal and likely due to rounding, the table is considered reproducible.
- **Table A7 Reproduced.** Verified against sheets A7_A and A7_B of `xml_7-21-2025.xls`. Some formatting and labeling differences were identified, but all numerical values are identical; therefore, the table is considered reproducible.

- **Table A8 Reproduced.** Verified against sheets A8_A and A8_B of xml_7-21-2025.xls. Minor formatting and labeling differences were identified, along with a change in the variable ordering; however, all numerical values are identical. Therefore, the table is considered reproducible.
- **Table A9 Reproduced.** Verified against sheets A9_A and A9_B of xml_7-21-2025.xls. Some formatting and labeling differences were identified, but all numerical values are identical; therefore, the table is considered reproducible.
- **Table A13 Reproduced.** Verified against sheets A13_A and A13_B of xml_7-21-2025.xls. Minor formatting and labeling differences were identified, along with a change in the variable ordering; however, all numerical values are identical. Therefore, the table is considered reproducible.
- **Table A14 Reproduced.** Verified against the estimates displayed in the Stata results window.

Reproduction Environment

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

- OS: Windows 11 Enterprise
- Processor: Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz
- Memory available: 16.0 GB
- Software version: Stata 19.5 MP, R 4.5.1