



## *Tax Compliance in Romania*

*Second Submission: RR\_ROU\_2023\_100*

*María Reyes Retana*

*reproducibility@worldbank.org*

*March 28th, 2024*

This review verifies the reproducibility of the exhibits included in the paper “*Tax Compliance in Romania*”.

### **Contents in this review:**

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

### *Main findings*

- This verification process contains two distinct stages, necessitated by the usage of confidential sources. While the replicators were granted access to a subset of the intermediate EU data, the full process could not be fully run. Consequently, we conducted a virtual reproducibility verification session on March 26th, 2024. During this session, the authors executed the initial segment of the code that generated the intermediate data. This intermediate dataset was subsequently provided to the replicator for further verification of this paper. It is pertinent to note that no portion of the Tax data could be shared with the replicators due to its confidential nature. Instead, the authors provided exhaustive log files, enabling us to validate the reproducibility of their findings. This report delineates our findings by combining both stages: the portion directly reproduced by the authors using the intermediate data and the analysis based on the comparison of log files.
- The code from intermediate data was successfully executed on a new computer after:
  1. Changing the file path.
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code running from intermediate data takes approximately 3 minutes.
- We conducted our reproducibility analysis based on the paper shared by the authors in their reproducibility package.
- Every exhibit was reproduced with accuracy.
- **Reproducibility Summary:**
  - **Data:** All data is confidential and not included in the package (details provided in README).

- **Code:** All code files are included in the package.
- **Outputs:** All outputs are generated by code included in the reproducibility package.
- **Reproducibility verification:** Reviewers used data provided directly by the authors to conduct the reproducibility verification, and this is not included in the package.

### *List of exhibits and reproducibility status*

#### **Results in the Main Section of the Paper**

- **Table 1** Does not show analysis results.
- **Table 2** Does not show analysis results.
- **Table 3 Reproduced.** The replicators utilized the code to accurately replicate the results for columns 2 and 4. However, the verification of data for column 3 was impeded by the lack of access to the file `Data/Tax_07_2020_for_imputation.dta`. Despite this, through virtual verification and reviewing the log files provided by the team, we managed to confirm the numbers for the third column, specifically. Line: `sum pyo1og, detail`.
- **Figure 1 Reproduced.** The visual appearance differs slightly: the output produced by the code features dotted lines, whereas the figure depicted in the document has a red solid line. These aesthetic differences do not compromise the reproducibility.
- **Table 4 Reproduced.** Direct verification was feasible solely for columns 2 and 4, where the data were accurately confirmed. Regarding column 3, verification was achieved through analysis of the log files supplied by the authors, specifically examining the lines related to:
  - `inequal7 vbrut if sample==1`
  - `pshare vbrut if sample==1, percentiles (50 90)`
  - `pshare vbrut if sample==1, percentiles (99)`
- **Table 5 Reproduced but requires manual steps.** We successfully verified all figures within the Tax Compliance column. This was achieved by dividing the values at the 50th percentile of the tables generated by the code, as displayed in the console, by the 50th percentile values of subsequent tables, in alignment with the methodology shown in the master Excel file. The Share of Employees metric was also verified directly produced by the code.
- **Table 6 Reproduced but required manual steps.** The table was successfully reproduced by executing the specific lines of code detailed in the "figures and tables do" file. The procedure involved taking the 50th percentile value generated by the first line of code (Median Survey Income) and dividing it by the value produced by the second line of code (Median Imputed Tax Income). Although I was able to deduce this method using your master check file, it would be beneficial for future clarity and reproducibility to include these steps in the README file. (code lines 50-116)
- **Table 7 Reproduced but required manual steps.** This table was successfully reproduced by following the specific lines of code outlined in the "figures and tables do" file, using the same procedure as the figure previously mentioned.

- **Table 8 Reproduced.** The verification of the second and third columns was successfully completed. The data for the third column were confirmed through analysis of the log files provided by the authors, specifically examining the Stata command line: `(tab mw_status if work_schedule==1)`.

### *Reproduction Environment*

- Paper exhibits were reproduced in a computer with the following specifications:
  - OS: Windows 11 Enterprise
  - Processor: Intel(R) Core(TM) i5-1145G7 CPU @ 2.60GHz
  - Memory available: 15.7 GB
  - Software version: Stata version 18