Reproducible Research Repository

Which Data Do Economists Use to Study Corruption? A Cross-Section of Corruption Research

Third Submission: RR_WLD_2025_274
Xuerong Lei, Mahin Tariq
reproducibility@worldbank.org
March 4, 2025

This review verifies the reproducibility of the exhibits included in the paper *"Which Data Do Economists Use to Study Corruption? A Cross-Section of Corruption Research"*.

Contents in this review:

1. Main findings

2. List of exhibits and reproducibility status

3. Reproduction Environment

Main findings

- **Every exhibit has been reproduced accurately.**

- The code was successfully executed on a new computer after:

  1. Change the file paths in the main do file

- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.

- The code takes approximately a minute to run.

- We conducted our reproducibility analysis based on the paper shared by the authors in the reproducibility package on March 3rd, 2025.

- **Reproducibility Summary**:

  - **Data**: All data sources are publicly available and included in the reproducibility package.
  - **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.
  - **Reproducibility verification**: Reviewers had access to the same materials in the public reproducibility package. The reviewers did not verify if publicly available data matches the data in the reproducibility package.
  - **Dependencies environment**: The reproducibility package does not use external dependencies.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Table 1** Reproduced Values for "Total" are missing from the code output but the sum of columns match the paper so the table is considered reproducible.

- **Table 2** Reproduced.

- **Table 3** Reproduced.

- **Table 4** Reproduced Values for "Total" are missing from the code output but the sum of columns match the paper so the table is considered reproducible.

- **Table 5** Reproduced.

- **Table 6** Reproduced.

- **Table 7** Reproduced.

- **Table 8** Reproduced. Values for "Total" are missing from the code output but the sum of columns match the paper so the table is considered reproducible.

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 18.0 MP