

README for Reproducing Results

Reproducible Package: “Lying to the Taxman or Accepting a Helping Hand? Evidence from a Novel Experiment on SMEs in Tanzania” Overview

This **README** file provides step-by-step instructions for reproducing the results presented in our paper. The folder structure is organized to streamline the process, with specific directories for outputs, data, and code. Please follow the instructions carefully to ensure the accurate replication of our results.

Folder Structure

1. **admin**: Folder to reproduce analysis from administrative data (Tanzania Revenue Authority).
 - **code**: R and STATA code to reproduce the analysis
 - **data-clean**: Contains fully processed data
 - **data-raw**: Contains the original, unprocessed data
 - **renv**: Create a reproducible environment for the project
 - **tables**: Contains the tables generated from analyzing administrative data
2. **survey**: Folder to reproduce analysis from the two survey rounds
 - **code**: R and STATA code to reproduce the analysis
 - **data**:
 - **analysis**: Contains fully processed data
 - **raw**: Contains the original, unprocessed data
 - **renv**: Create a reproducible environment for the project
 - **results**: Contains results (tables) generated from analyzing survey data

Prerequisites

To successfully run this reproducible package, the following software must be installed on your computer:

- R
- STATA

Steps to Reproduce Results - Survey

1. Clone or download this repository to your local machine.
2. Open the **code/** folder in the replication package.
3. Run 00_master.do to obtain all the output from the first survey round at once (after changing the path).
4. Alternatively, run the Stata and R code in the following order (don't forget to change the path):
 - 01_data_prep.do
 - 02_regressions.do
 - tax_project_tables_21jun.R (initiate from the project file)
5. The output generated by the analysis will be saved in the **results/** folder.
6. Open the **tables/** folder and run Table* files to produce the tables.

Steps to Reproduce Results – Admin

1. Open the **code/** folder in the replication package, then the **STATA/** folder.
2. Run 01_split_tra.do.
3. Initiate the project by selecting admin_analysis.Rproj.
4. Open **code/** folder in RStudio, and then open the **R/** folder.
5. Follow the sequencing of the code to generate outputs stage by stage.
6. Return to the **code/** folder, open the **STATA/** folder, and run 06_regressions.do and 07_census.do.

7. See the output in the **tables/** folder (Table 1 is obtained from the Tanzania Revenue Authority website).

Data Availability Statement

Data Sources and Access

1. Monthly Tax Payment Data from the Tanzania Revenue Authority

- Dataset name: 202201_202403_Revo
- Includes monthly tax payment data from January 2022 to March 2024.
- These data are private and shared by the Tanzania Revenue Authority under agreement.
- Cannot be disclosed (for access reach out to the Tanzania Revenue Authority).
- Data shared with us was strictly anonymized.
- This dataset should be placed in the *admin/data-raw* folder.

2. List of Subnational Divisions

- Dataset name: GN Latest
- A list of subnational divisions.
- Cannot be disclosed (for access reach out to the National Bureau of Statistics).
- This dataset should be placed in the *admin/data-raw* folder.

3. Survey Data (First Round Interviews)

- Dataset name: Project Biashara Final Data set 23122022.2
- Survey data from the first round of interviews (December 2021).
- This file is included in the data folder within this package and is publicly available.
- This dataset should be placed in the *survey/data/raw* folder.

4. Survey Data (Second Round Interviews)

- Dataset name: tanzania_followup_may2024_v7
- Survey data from the second round of interviews (May 2024).
- This file is included in the data folder within this package and is publicly available.

- This dataset should be placed in the *survey/data/raw* folder.
- 5. Tanzania's Population and Housing Census (2022)**
- Dataset name: phc2022
 - Tanzania's Population and Housing Census (2022). A 10% subsample of Tanzania PHC 2022 can be accessed through this website:
<https://microdata.nbs.go.tz/index.php/home>
 - This dataset should be placed in the *admin/data-clean* folder.

Restrictions and Terms of Use

- **202201_202403_Revo**: Due to the confidentiality agreements with the Tanzania Revenue Authority, these datasets cannot be shared publicly. Access is restricted and requires authorization from the Tanzania Revenue Authority.

Troubleshooting

If you encounter any issues, please ensure that:

- Both R and STATA are correctly installed and operational on your machine.
- Directory paths are correctly specified in both the R and STATA scripts.
- STATA scripts are executed in the correct order if run manually.

Conclusion

By following these steps, you should be able to accurately reproduce the results from our paper.

Contact

If you have any questions or issues with the replication package, please contact the authors: Christopher Hoy, Jonathan Karver, and Revocatus Paul at rpaul4@worldbank.org.