README for Reproducing Results

Reproducible Package: "Trade-offs in the Design of Simplified Tax Regimes: Evidence from Sub-Saharan Africa"

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. **Outputs**:
    - **Figures**: Contains all the figures generated for the paper.
    - **Tables**: Contains all the tables generated for the paper.
2. **Data**:
    - **Raw**: Contains the original, unprocessed data.
    - **Intermediate**: Contains data that has been partially processed.
    - **Final**: Contains fully processed data ready for analysis.
        - **Experiment**: Sub-folder with data from:
            * **Scale Up**: Data from the scaled-up experiment.
            * **Pilot**: Data from the pilot study. Note: The pilot data underwent multiple iterations and slight modifications in collaboration with the Kenya Revenue Authority.
3. **Code**:
    - **Master Script**: The main R script that runs all other scripts. Adjust directories in this script before running.
    - **Randomization Scripts**: Written in STATA. These scripts are numbered and can be run through R or manually if issues arise.

Prerequisites

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

- **R**
- **STATA**

Steps to Reproduce Results

1. **Setup Directories**:
    - Open the **Master Script** in the **Code** folder.
    - Modify the directory paths in the script to reflect the location of the data and output folders on your computer.
    - Similarly, open the `.do` files in the STATA scripts and update the directory paths accordingly.
2. **Running the Master Script**:
    - Run the Master Script in R. This script will sequentially execute all necessary scripts, including data processing and analysis.
    - If the script encounters issues running the STATA randomization scripts through R, follow the numbering and run these scripts manually in STATA.
3. **Manual Execution of STATA Scripts**:
    - Navigate to the **Code** folder.
    - Identify the numbered STATA scripts.
    - Open each script in STATA and run them in the specified order.

Special Considerations

- **Pilot Data Randomization**:
  - The randomization process for the pilot data slightly varied due to iterations and changes in collaboration with the Kenya Revenue Authority. Ensure that the steps in the STATA scripts reflect these modifications accurately.
- **Scale Up Randomization**:
  - The scale-up data was randomized using the updated dataset. Follow the scripts for precise steps.

Data Availability Statement

Data Sources and Access

1. **Three batches of ToT data**:
   - Includes both ToT filing and payments.
   - These data are private and shared by the Kenya Revenue Authority under agreement.
   - Cannot be disclosed.
   - Data shared with us were strictly anonymized.
2. **custom.geo.json**:
   - A standard JSON file used for producing a map.
   - It can be downloaded here.
3. **str_overview.xls**:
   - A table generated by collecting data from multiple sources by the authors.
   - This file is included in the data folder.
4. **VAT.csv**:
   - Private data shared with us with VAT collection transactions.
   - Cannot be disclosed.

Restrictions and Terms of Use

- **ToT Data and VAT Data**: Due to the confidentiality agreements with the Kenya Revenue Authority, these datasets cannot be shared publicly. Access is restricted and requires authorization from the Kenya Revenue Authority.
- **custom.geo.json**: Available for download and use without restrictions.
- **str_overview.xls**: Created by the authors and can be used for replication of results as described in this document.

Contact Information

For any inquiries regarding data access, please contact the corresponding author or the Kenya Revenue Authority for data that cannot be publicly disclosed.

Generating Figures and Tables

- **Figures 1 to 13**: Generated using the script `2_graphs.R`.
- **Tables 1, 2, and 4**: Generated using the script `3_tables.R`.
- **Table 3**: Generated using the script `7_table_main_results.R`.
- **Tables 5 and 6, Figures 16 and 17**: Generated using the script `8_other_scale_outputs.R`.

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. If you have any questions or require further assistance, please contact rdoino@worldbank.org .