README and Guidance

for

How Can Lower-Income Countries Collect More Taxes? The Role of Technology, Tax Agents, and Politics

Oyebola Okunogbe and Gabriel Tourek[†]

November 17, 2023

[*]Prepared for *Journal of Economic Perspectives*.

[†]Affiliations: World Bank (ookunogbe@worldbank.org), University of Pittsburgh (gabriel.tourek@pitt.edu).

Contents

1 Overview

The code in this replication package constructs the analysis datasets for "How Can Lower-Income Countries Collect More Taxes? The Role of Technology, Tax Agents, and Politics" using Stata and R. The Stata master code file 0_Master.do runs all code to generate the data for Figures 1–4 and Appendix Table 1 in the Online Data Appendix. The R code file Figure5_Replication.R produces Figure 5. Replicators can expect the code to run in 30 minutes.

2 Data Availability and Provenance Statements

2.1 Statement about Rights

We certify that the authors have legitimate access to and permission to use the data used in this manuscript.

2.2 Statement of Data Availability

All data is obtained from public sources.

2.3 Details on Data Sources

The paper uses publicly-available data obtained from the UNU-WIDER Government Revenue Dataset (UNU-WIDER, 2022a) and UNU-WIDER World Income Inequality Database (UNU-WIDER, 2022b) (Figures 1 and 2, Appendix Table 1), the International Survey on Revenue Administration (CIAT-IMF-IOTA-OECD, 2022) and OECD Inventory of Tax Technology Initiatives (ITTI) (OECD, 2023) Global Survey on Digitalisation (Figures 2 and 3, Appendix Table 1), Lagos State Government statistical digests (Lagos State Government, 2006, 2010, 2013) (Figure 4), and details of existing studies compiled by the authors (Figure 5). Copies of all data used by the authors can be downloaded from the American Economic Association repository accessible here: http://doi.org/10.3886/E194822V1 (Okunogbe and Tourek, 2023).

All data we use in publicly available. Copies of these data are included as part of the replication package, and the data can be downloaded directly via the links provided above.

1. UNU-WIDER Government Revenue Dataset (2022): Contains country-level measures of tax variables. The dataset copy used by the authors is included in the replication directory in Data/01_base/unuwider_data. More details can be found here: https://doi.org/10.35188/UNU-WIDER/GRD-2022

2. UNU-WIDER World Income Inequality Database (2022): Contains country-level measures of inequality. The dataset copy used by the authors is included in the replication directory in Data/01_base/unuwider_data. More details can be found here: https://doi.org/10.35188/UNU-WIDER/WIID-300622

3. International Survey on Revenue Administration (CIAT, IMF, IOTA, OECD 2022): Contains survey data collected from national tax authorities. The dataset copies used by the authors is included in the replication directory in Data/01_base/rafit_data. More details can be found here: https://www.imf.org/en/Capacity-Development/Training/ICDTC/Courses/ISORA

4. OECD Inventory of Tax Technology Initiatives (ITTI) (2023) Global Survey on Digitalisation: Contains inventory of tools and digitization solutions implemented by national tax authorities. The dataset copies used by the authors is included in the replication directory in Data/01_base/taxtech_data. More details can be found here: https://www.oecd.org/tax/forum-on-tax-administration/tax-technology-tools-and-digital-solutions/

5. Lagos State Government Digest of Statistics (2006, 2010, 2013): Contains measures of revenues for Lagos State, Nigeria. The dataset manually entered from the gazettes by the authors is included in the replication directory in Data/01_base/lagos_data.

6. Details of existing studies: Contains details of interventions and measures of effect sizes as described in existing studies of taxation. These details were compiled from individual publications by the authors. The dataset compiled by the authors is included in the replication directory in Data/01_base/intervention_data.

3 Dataset List

The table on the following page lists the base datasets included in the replication directory or produced by replication code. All datasets are located in or saved to the "Data" folder of the replication directory.

Data File	Source	Notes	Provided
Data/01_base/unuwider_data/UNUWIDERGRD_2022_Merged (Oct).dta	UNU-WIDER	Public	Yes
Data/01_base/unuwider_data/WIID_30JUN2022_0.dta	UNU-WIDER	Public	Yes
Data/01_base/rafit_data/1_–12_.xlsx	CIAT, IMF, IOTA, OECD	Public	Yes
Data/01_base/taxtech_data/DI1–TT7.xlsx	OECD	Public	Yes
Data/01_base/lagos_data/Lagos_total_tax_99_17.xlsx	Lagos Govt.	Public	Yes
Data/01_base/intervention_data/Intervention_Database.xlsx	Studies	Public	Yes

All data sources are in one of the following formats: .dta or .xlsx.

4 Computational Requirements

4.1 Software Requirements

Stata (code was last run with Stata/SE version 18 for Mac), with additional packages:

- estout
- grstyle
- palettes
- revrs
- blindschemes

The replication package includes all the files of these additional packages and no installation of Stata dependencies is required. Line 24 of program "0_Master_.do" points out the correct location of these files for Stata to load them.

R code was last run with version 4.2.0. The replication package includes environment files created with the R package `renv` that install the correct versions of all R dependencies in lines 7-8 of program "Figure5_Replication.R".

4.2 Memory and Runtime Requirements

Reproducing the analysis requires approximately half and hour on a standard desktop machine with Stata version 18 or higher and R version 4.2.0 or higher.

The code was last run on a 2 GHz Intel® Core™ i5, 8 GB 1867 MHz LPDDR3, on macOS Sierra Version 10.12.6.

The guidelines provided below outline the computational requirements for Stata/MP and Stata/SE (sourced from `https://www.stata.com/products/compatible-operating-systems/`):

5 Description of Programs and Code

The following code files are located in the Code folder of replication directory.

- 0_Master.do will run the entire Stata replication for Figures 1–4 and Appendix Table 1 from start to finish, setting dependencies and calling all other programs

in the replication directory

- 1_Package_Setup.do will set graph styles necessary to reproduce the plots
- 2_Data_Construction.do will clean and append datasets to produce intermediate datasets and analysis datasets used to produce Stata output
- 3_Main_Tables_Figures.do will create Figures 1–4 and Appendix Table 1 in the main paper, calling on individual dofiles in Code/Tables_Figures
- Figure5_Replication.R will create Figure 5

6 Instructions for Replicators

To run entire replication:

1. Enter default path in Code/0_Master.do, line 23
2. Run Code/0_Master.do to run all code files in sequence
3. Open JEP-tax-2023.Rproj in RStudio
4. Open Code/Tables_Figures/Figure5_Replication.R in the same RStudio session and enter the correct path in line 22
5. Run Code/Tables_Figures/Figure5_Replication.R. Line 7 should be activated if the user does not have `renv` installed and line 8 only needs to run once in a computing environment to install packages
6. Compile outputs from Output folder

If running files individually:

- 0_Master.do lines 1–34 must be run at the beginning of each session
- Code files to clean and combine datasets (2_Data_Construction) must be run before any used to produce paper exhibits (those listed in Section 7)

7 List of Paper Exhibits and Programs

Code files listed below reproduce paper tables and figures. Note that descriptive text added to paper tables and figures — e.g., fixed effects and controls included in regressions — may not appear in replication output but are observable in the code that produces specific outputs.

Exhibit	**Program**	**Line(s)**	**Output Files**
Figure 1	Figure1_Replication.do	133	Figure1.pdf
Figure 2	Figure2_Table1_Replication.do	89	Figure2.pdf
Appendix Table 1	Figure3_Replication.do	124	Table1.tex
Figure 3	Figure3_Replication.do	120	Figure3.pdf
Figure 4	Figure4_Replication.do	13	Figure4.pdf
Figure 5a	Figure5_Replication.R	76	Figure5a_compliance.png
Figure 5b	Figure5_Replication.R	135	Figure5b_revenues.png

References

CIAT-IMF-IOTA-OECD, "International Survey on Revenue Administration: 2020 and 2021," https://data.rafit.org/?sk=8b008788-ebde-4d61-bc90-7438d6aa12dc&sId=1637191076670 2022.

Lagos State Government, "Digest of Statistics," https://mepb.lagosstate.gov.ng/wp-content/uploads/sites/29/2017/08/Digest-of-Statistics-2006.pdf 2006.

_ , "Digest of Statistics," https://lagosstate.gov.ng/wp-content/uploads/sites/29/2017/01/2010-Digest-of-Statistics.pdf 2010.

_ , "Digest of Statistics," https://mepb.lagosstate.gov.ng/wp-content/uploads/sites/29/2017/01/2013-Digest-of-Statistics.pdf 2013.

OECD, "OECD Inventory of Tax Technology Initiatives," https://www.oecd.org/tax/forum-on-tax-administration/tax-technology-tools-and-digital-solutions/ 2023.

Okunogbe, Oyebola and Gabriel Tourek, "Replication Data for: How Can Lower-Income Countries Collect More Taxes? The Role of Technology, Tax Agents, and Politics," Technical Report, American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor] 2023.

UNU-WIDER, "Government Revenue Dataset," https://doi.org/10.35188/UNU-WIDER/GRD-2022 2022.

_ , "World Income Inequality Database (WIID)," https://doi.org/10.35188/UNU-WIDER/WIID-300622 2022.