

Data and Code for “Technology, Taxation, and Corruption: Evidence from the Introduction of Electronic Tax Filing”

Oyebola Okunogbe and Victor Pouliquen (victorpouliquen@gmail.com)

Data Availability Statement

We obtained the taxpayer-level data used in the paper directly from the Tajikistan Tax Committee, for the purpose of conducting this research, and subject to the condition that no confidential information would be shared with outsiders. As the Tax Committee does not currently have a formal policy for entering data use agreements with researchers, anyone wishing to obtain access to the data would need to approach the senior management at the Tax Committee and make a business case to them for data access. We would be happy to run any sensitivity analyses ourselves and share the results with anyone wishing to run such analyses.

Requirements

1. Software requirements:

Stata, R, Latex, Excel

The following Stata commands are required (installed automatically when the dofile is run):

cdfplot, outreg2

2. Time requirements:

The R codes can take a few hours to run.

Steps to follow in order to replicate all Tables and Figures in the paper

Step 1: TABLES 1 to A11 and Figures 1 to A2

* TABLES 1 to A11 and figures 1 to A2 are created using Stata with the following steps:

1- Run the DoFile: TJ_analysis (In the folder "2_dofiles")

- This dofile opens the Raw data, creates the variables used in the analysis and saves the database used for the analysis

- To run the dofile, you'll need to change the directories in the globals at the beginning of the dofile

2- Run the DoFile: TJ_tables_creation

- This dofile creates all the Tables and Figures in the paper.
- Tables are exported in .out format and figures in .png.
- To run the dofile, you'll need to change the directories in the globals at the beginning of the dofile

3- Open the Excel File "Efiling_paper_MASTER_Tables_and_Figures"

- This File includes all the tables and figures in the paper.
- Results in the Excel file are linked to the .out files created by the dofile.
- When the ".out" files are open (from the master Excel file) the results in the master file are updated.

Step 2: TABLES A12, A13 and A14

TABLES A12, A13 and A14 are created using R with the following steps:

1- Run the R code "ML_Tables_TJpaper" located in the folder ("4_ML_analysis").

- To run the R code, you'll need to change the directories at the beginning of the code.
- This R code uses another R code located in the same folder ("ML_Function").
- This code uses the raw database in the same folder ("TJ_data_for_ML.dta") and performs Machine learning analysis following Chernozhukov, Demirer, Duflo and Fernandez-Val (2018)
- .txt files are created as well as figures (that we don't use in the paper).
- Note that this code can take a few hours to run.

2- Using LATEX: run the .tex file "TJ_ML_import_tables".

- This latex code creates tables from the .txt files created by the R code.

- For the tables A13 and A14, the R and Latex codes create tables showing the results for the two most effective machine learning methods (in our case Random Forest and Boosting) as in Chernozhukov et al. (2018)
- In the paper, we focus on the Random Forest for simplicity but the two methods give similar results.

3- Finally, we pasted results from Table A12, A13 and A14 in the Latex file to the master Excel File "Efiling_paper_MASTER_Tables_and_Figures"

Important note:

- Results for tables A12, A13 and A14 might be slightly different from the results in the paper when the code is run on a different computer.
- This is due to a random component included in the machine learning code provided by Chernozhukov, Demirer, Duflo and Fernandez-Val (2018).
- While the exact numbers are different, the results are always very close and qualitatively identical.

Step 3: Figure A3

Figure A3 is created using R with the following steps:

1- Run the R code "grf_var_imp_TJpaper" located in the folder ("4_ML_analysis").

- To run the R code, you'll need to change the directories at the beginning of the code.
- This code creates the Figure A3 "GRF_varimport_index_TJ_Efiling"

2- Paste the figure in the Master Excel file "Efiling_paper_MASTER_Tables_and_Figures".

- Variable labels are added manually in the Excel file.