Tax Compliance and Gender: Evidence from Field Experiment In Ethiopia

Alemayehu Ambel and Firew Bekele Woldeyes

March 2026

1 Overview

The repository package includes data, codes, and output from the paper "Tax Compliance and Gender: Evidence from Field Experiment in Ethiopia". This document provides an overview of the data, computational requirements, programs, and instructions to reproduce the results of our research and is structured as follows. Section 2 describes the data source and ownership and how it can be accessed. Section 3 describes the data set used in the analysis. Section 4 provides details on computational requirements. Section 5 provides instruction to replicators.

2 Data Availability and Provenance Statements

2.1 Statement about Rights

Two sources of data are used in the analysis presented in the paper. The first data set is obtained from the Ethiopian Customs and Revenue Authority, which is now the Ministry of Revenue. The Ministry of Revenue Retains the ownership of the anonymized tax data.

2.2 Statement of Data Availability

File Name	Electronic Sales Registry Machine Utilization
Source	Ministry of Revenue
URL	Not Publicly Available
Access Date	2018
Citation	Ministry of Revenue (2018)

File Name	Firm Characteristics
Source	Ministry of Revenue
URL	Not Publicly Available
Access Date	2018
Citation	Ministry of Revenue (2018)

File Name	Financial Statements
Source	Ministry of Revenue
URL	Not Publicly Available
Access Date	2018
Citation	Ministry of Revenue (2018)

File Name	Taxpayer Category
Source	Ministry of Revenue
URL	Not Publicly Available
Access Date	2018
Citation	Ministry of Revenue (2018)

File Name	Profit Tax Paid
Source	Ministry of Revenue
URL	Not Publicly Available
Access Date	2018
Citation	Ministry of Revenue (2018)

License: The right to access and use anonymized data on tax declarations and payments in Ethiopia is restricted. Although the Ethiopian Customs Revenue Authority (now the Ministry of Revenue) granted access to the data for academic purposes in 2013, 2015, 2017, and 2018, individual assessments are required, and public sharing of the data is prohibited.

For inquiries, the Ministry of Revenue can be contacted through the following channels: Email: info.mor@mor.gov.et Phone: +251 11 662 98 00 P.O. Box: 2559 Addis Ababa, Ethiopia

File Name	Taxpayers' treatments status
Source	[Shimeles et al.(2017)Shimeles, Gurara and Woldeyes]
URL	Not Publicly Available
Access Date	2015
Citation	[Shimeles et al.(2017)Shimeles, Gurara and Woldeyes]

The second dataset used in the paper includes the treatment groups and gender of the taxpayers from [Shimeles et al.(2017)Shimeles, Gurara and Woldeyes].

License: [Shimeles et al.(2017)Shimeles, Gurara and Woldeyes].

I am one of the co-authors of the paper.

2.3 Details on Data sources

A detailed description of the data sources, including collection methods and any relevant transformations, is provided in Appendix A.

3 Dataset and file structure

All files in the Reproducibility package:

- The data folder contains the data.
  - The data directory contains two subfolders: 'input' and 'output'. Data from six modules are located within the 'input' subfolder. The 'output' folder is currently empty; however, after executing the data preparation do file, it will house the final dataset for analysis.
- The output directory contains two subfolders: one for figures and one for tables.
- The dofiles folder contains all do files, there is a folder labeled do with two subfolders: analysis and data preparation.

4 Computational Requirements

Stata is used for the analysis in this project. This project is written in Stata 16 MP and has not been tested on older versions. The codes run on: Operating system: Windows 11. Processor: 13th Gen Intel(R) Core(TM) i7-1365U 1.80 GHz Installed RAM: 16.0 GB (15.7 GB usable)

4.1 Software Requirements

4.2 Packages

Packages installed from ssc are provided in the main.do file.

5 Instructions for Replicators

Data shared with the reproducibility team are artificial datasets produced via random number generation. Although the data values are fictitious, the format and structure remain identical to the source data. These datasets were generated at the raw level as the actual data. Therefore, the same merging and transformation processes were used as it would be the case for actual data preparation to ensure reproducibility.

5.1 Data preparation

0master_taxation_gender is the main do file for data_preparation. It sequentially runs the following do files:

1. 1import_fs: imports financial statement.
2. 2gender_char: imports taxpayer info, including gender and treatment.
3. 3cat_char: imports categories of taxpayers.
4. 4fiscal_device: imports fiscal device usage data.
5. 5profittaxpaid: imports data on the profit tax paid.
6. 6firm_chara: imports firm characteristics like sector and location.
7. 7merge_gender_cat_fin_stat: merges files using appropriate IDs and year.
8. 8cleaningandpreparation: cleans, labels, and prepares data for analysis.

After running 0master_taxation_gender, an output .dta file is saved in the data\otuput folder, ready for analysis. To run the do file, changing the path in the first part of the code is required.

5.2 Analysis

- Running main.do will produce all the results in the paper. To run the main.do it is required to change the path in part 0 at the top.
- Do files are structured in a way that the output from the descriptive regression analysis is in separate files.
- All do files will run by the main.do. The tables are exported from stata to LaTeX file and stored in the output.
- 1.1 Analysis Descriptive.do produces Table 1 to Table 3.
- 1.2 Analysis Regression.do reproduces the regression results reported in Table 4 to 6. The results are stored in the folder 'outputs / tables', which is clearly labeled with the table number.

6 Appendix

6.1 Data Sources

The data used for this document come from the Ethiopian Revenue and Customs Authority, now under the Ministry of Revenue. This comprehensive administrative database provides a diverse array of information, including declared sales, costs of goods sold, expenses, profits, tax obligations, deductions for losses, and payable profit tax. In particular, the administrative data source does not explicitly denote the gender of business owners. Gender categorizations were established by merging these administrative data with survey data obtained through a field experiment conducted by [Shimeles et al.(2017)Shimeles, Gurara and Woldeyes]. "Additionally, the treatment intervention is based on the same study.. Their study targeted 5,400 randomly selected taxpayers out of approximately 86,000 in Addis Ababa for fiscal year 2013/2014. The sampling method used was a stratified random sampling technique based on geographic location (referred to as sub-city) and economic sectors.

6.2 Data Transformation

In processing the data, we used the natural logarithm of monetary values, such as sales, tax declarations, payments, cost of goods sold, and expenses. Given the presence of zero values, we added one to the monetary values before applying the logarithmic transformation.

References

[Shimeles et al.(2017)Shimeles, Gurara and Woldeyes] Shimeles, A., Gurara, D.Z., Woldeyes, F., 2017. Taxman's Dilemma: Coercion or Persuasion? Evidence from a Randomized Field Experiment in Ethiopia. The American Economic Review 107, 420–424. URL: https://www.jstor.org/stable/44250433. publisher: American Economic Association.