United Nations
University
WIDER
World Institute for Development
Economics Research

national treasury
Department:
National Treasury
REPUBLIC OF SOUTH AFRICA

Guide to Creating Dummy Data for Datasets at the National Treasury Secure Data Facility

National Treasury Secure Data Facility (NT-SDF) dummy data

June 2024

Welcome!

This guide is designed to help researchers understand the datasets available at the National Treasury Secure Data Facility (NTSDF) and prepare their analysis scripts (do-files) before accessing the actual data. The current code is in its beta version, and we encourage researchers to provide feedback on their experience and suggestions for improvement. The beta version of the code generates the following datasets: Firm-level Corporate Income Tax (CIT) data, SARS-Treasury Matched Firm Panel (STMFP), Job-level IRP5 data, Return-level Value Added Tax (VAT) data, and Transaction-level Customs data.
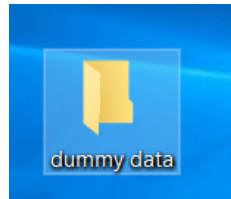
Important Notes:

- The dummy data does not mirror the exact structure or relationships between variables found in the real datasets. All numeric variables in the dummy data are populated with completely random numbers.

- The dummy data code is compatible with STATA versions 17 and 18. Please ensure you use one of these versions to execute the code.

- Researchers should note that the latest version of STATA available at the NTSDF is version 18. Therefore, any updates or features added to newer versions of STATA will not be supported at the NTSDF.

- For any questions about the code, please reach out to ntsdf@treasury.gov.za

- The code is designed to run smoothly on both Windows and Mac if the following instructions are followed. Any modifications to the code may hinder its performance.
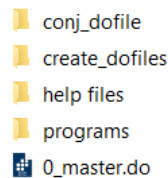
Instructions:

1. **Create a Project Folder:**

- On your computer, create a new folder named "dummy data". This will be your project workspace.
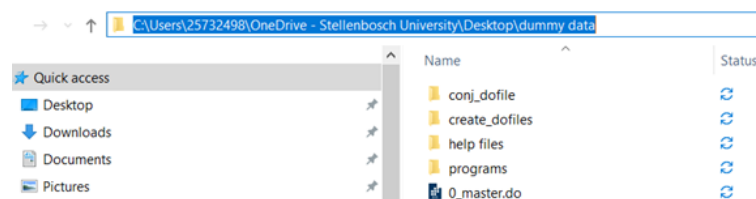

2. **Download and Extract Files:**

   - Download the zipped folder "dummy data files" and extract its contents. This folder contains a file named "0_master_do", as well as subfolders like "conj_dofile", "create_dofiles", and "programs".

3. **Copy Files:**

   - Copy all the extracted files from step 2 into your "dummy data" folder created in step 1.
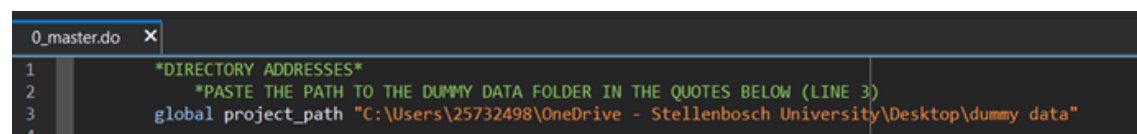

4. **Open Stata File:**

   - Open the Stata file "0_master.do".

5. **Update Global Variable:**

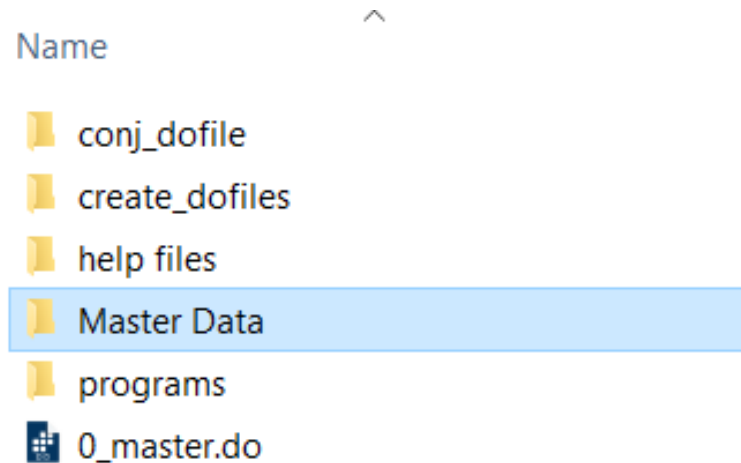   - Update the "project_path" global variable by pasting the path of your "dummy data" folder.

2

United Nations University WIDER
World Institute for Development Economics Research

national treasury
Department:
National Treasury
REPUBLIC OF SOUTH AFRICA

6. **Run Master Do-file:**

   - Execute the "0_master.do" file.
   - Expected runtime: Approximately 50 minutes

7. **Access Dummy Datasets:**

   - The code generates a "Master data" file. Once the "0_master.do" script has finished running, the dummy datasets will be available in the Master data folder

Name

conj_dofile

create_dofiles

help files

Master Data

programs

0_master.do

By following these steps, you will generate dummy data that aids in familiarizing yourself with the datasets and preparing for your analysis. If you have any questions or encounter any issues, please don't hesitate to contact the NTSDF data team (ntsdf@treasury.gov.za) for assistance.