README

This README file provides instructions to replicate the outputs in "Multiproduct-Exporters: Verifying Stylized Facts", by Lena Sheveleva

It is divided into two main sections. In Part 1, I provide an overview of the Main.do file executed in Stata. This file generates Figures 1-2 and Tables 1-3 in the main text, along with Figures 1-10 and Tables 1-7 in the Appendix. Part 2 discusses the generation of Appendix Figures 11 A and B, achieved using Jupyter Notebook Python scripts.

Part 1

The **Main.do** file is a Stata script serving as the master file for replicating figures and tables from Sheveleva (2014). It constructs Figures 1-2, Tables 1-3 in the main text, and Figures 1-10, Tables 1-7 in the Appendix.

The Main.do file calls the following files that produce the corresponding inputs:

Do File	**Corresponding Output/Function**
PMC-world-data-permutation.do	Creates a dataset of permuted data
Table 3-5 - Appendix.do	Tables 3,4,5 in the Appendix
PMC-world-for-plots.do	Prepares permuted data plotting Figures 1 and 2 in the main paper and Figures 2-10 in the Appendix
PMC-within-country-data-permutation.do	Creates within country permutations of sales
Most_Poppular_Destinations.do	Identifies most popular export destinations
Figure1A.do	Creates Figure 1A in the main paper
Figure1B.do	Creates Figure 1B in the main paper
Table1.do	Creates Table 1 in the main paper
Tables-2-and-3.do	Creates Tables 2 and 3 in the main paper
Figure-1-Appendix.do	Creates Figure 1 in the Appendix
Table1-Appendix.do	Creates Table 1 in the Appendix
Table2-Appendix.do	Creates Table 2 in the Appendix
Tables-6-ADjR2.do	Creates Tables 6 in the Appendix
Table7-Appendix.do	Creates Table 7 in the Appendix

All of the files invoked in Main.do are availiable in folder Code-Stata.

Necessary Data Inputs

There are two key data inputs.

The first data file is **China.dta**, a sample of exports of the Chinese exporters and includes the following variables:

Variable Name	Description	Data Type
F	Firm identifier	String
D	Destination	String
Hs	Product code according to the 6-digit Harmonized System	String
Y	Year	String
V	Annual sales value for firm-product-destination	String

Second data file is **China_country_names.dta,** converts the **d** variable (destination) from the Chinese country classification to English country names and ISO codes (provided with the package).

The data files should be placed in Data-Stata folder. Users should ensure their dataset contain the variables above, properly renamed, and formatted according to the above specifications.

The dataset China.dta is in turn created using file **preprocessing.do** file also availiable in the Code-Stata folder. It uses raw data on the customs transactions data as well as the annual survey of industrial firms in 2003 that allows to match customs data with manufacturing firms. The results in this paper are robust to using Customs transactions only.

Note: The Annual Survey of Industrial Firms is collected by the National Bureau of Statistics of China. The customs transaction data is collected by the General Administration of Customs of China. They can be accessed at the discretion of China's National Bureau of Statistics and their authorized sellers and are not availiable to the public.

Further note, that the Main.do file can also replicate the results in this paper for Mexico using publicly availiable data from https://www.worldbank.org/en/research/brief/exporter-dynamics-database. The details of outputs for Mexico can be found https://github.com/LlenaChavel/Multi-product-Exporters-Verifying-Stylized-Facts. In this case, the Mexico data file should conform to the same requirements as the China.dta file.

Stata Version

This script was developed and tested with Stata 17. Users must ensure they have this version or a compatible version of Stata to successfully execute the script.

Execution Instructions

At the beginning of **Main.do**, global variables define the locations of project components:

Global Variable	Description
drive	The root location of the entire project which should contain folders:

	• Code-Stata holding the Main.do and all other relevant do files. • Data-Stata containing a country level exports dataset, such as China.dta or Mex.dta.
country	Identifies the country for which the analysis is carried out (e.g., China)
year	The year of the analysis (e.g., 2003)

Executing the code will create new directories in the **drive** folder:

- **wd**: a directory where intermediate data files are stored
- **Includes**: a directory where outputs are stored. The includes directory assigned the name Includes-country-year (e.g., Includes-China-2003)

To execute the script type, do **Main.do** in the Stata command window.

Outputs

The script generates Figures as pdf files and tables as tex files. As discussed above, the outputs are placed in the includes folder (Includes-China-2003). The tex files will require manual compiling to be converted into PDFs. This can be done with any Tex editor.

Troubleshooting

Should you encounter issues while running **Main.do**, verify the following:

- The dataset is correctly formatted with the appropriate variable names and types.
- Directory paths in the script are correctly set to match your system.
- Note, the individual do files called by the Main file cannot be run individually.
- Synchronizing files in the **$drive/wd** directory using Dropbox or One-Drive software may lead to write,availability or sharing errors.

Part 2

Executing AppendixFig11 Notebook:

The notebook "AppendixFig11" replicates Appendix Figures 11 A or B located in folder Code-python.

To replicate the Figures

- Set the **Fig** variable at the top of the script to "A" for Figure 11 A or "B" for Figure 11 B.
- Note, the code was written and run in Jupiter notebook (version 6.1.4) and may not work correctly in other python environments due to how Jupiter notebooks treat global and local variables.
- The program uses 60 threads to run in under 1 hour. Please adjust accordingly for your machine by changing 60 to the machine specific number:
  results = Parallel(n_jobs=60)(delayed(process_pair)(i) for i in pairs)
  (Note that reducing the number of threads may increase the execution time dramatically.)

Required Libraries:

- pandas: For data manipulation and analysis.
- numpy: For numerical operations.
- matplotlib.pyplot: For creating visualizations in Python.
- mpl_toolkits.mplot3d.Axes3D: For 3D plotting capabilities.
- plotnine: For creating complex plots using a grammar of graphics approach.
- statsmodels.formula.api: For fitting statistical models using formulas.
- numpy.random.default_rng: To create a new instance of the default random number generator.
- joblib.Parallel and joblib.delayed: For parallel computing, allowing for the execution of functions in parallel across multiple processors.
- multiprocessing.cpu_count: To determine the number of CPU cores available on the system.

Support

Replicators seeking assistance or wishing to report issues can contact the package author. (yas0306@gmail.com.)