

README for the Reproducibility Package for Policy Research Working Paper

This ReadMe file is created to guide future readers through the package and how to proceed with the replication of a working paper authored by Daniela M. Behr and Yue Sophie Xi, titled as “Closing the Gender Gap in Entrepreneurship: Overcoming Challenges in Law and Practice for Female Entrepreneurs”.

Contents

Data Availability Statement/Data Sources	1
Statement about Rights	2
Reproduction Instructions	2
Directories of Do-Files and List of Exhibits	3
Requirement.....	4
Software Requirement:.....	4
Folder Structure	5

Data Availability Statement/Data Sources

In this reproducibility package, we used the following data sources in “**datawork/input**” folder. All data used are publicly available:

- **WBL 1.0 Data for 1971 – 2024 (Panel Data):** a [publicly available dataset](#) on WBL’s website, covering WBL 1.0 data over time with annual data for 190 economies during the period from 1971 to 2024 (calendar years 1970 to 2023). Data were accessed on October 28, 2025. The file in the folder is:
 - “WBL1-0-1971-2024.dta”.
- **WBL 2.0 Data for 2024:** [publicly available datasets](#) on WBL 2.0 data encompasses the team’s newly revised questions assessing legal frameworks, supportive frameworks, and expert opinions. They cover the timeframe of WBL’s 2024 coding cycle (calendar year 2023). Data were accessed on October 29, 2025. The files in the folder are:
 - “WBL2024_LEGAL_EconLevel_ToPublish.dta”;
 - “WBL2024_SUPPORTIVE_EconLevel_ToPublish.dta”.

- **World Bank Official Boundaries:** Shapefiles of World Bank-approved administrative boundaries are downloaded from the [official website](#), last updated as of September 10, 2025. This is to make world map used in Figure 2. Shapefiles are stored in a sub-folder named “*datawork/input/World Bank Official Boundaries - Admin 0_all_layers*”.
 - The raw shapefile used is “*WB_GAD_ADM0_complete.shp*” and then stored as “*WB_GAD_ADM0_complete_shp.dta*” for use in Stata.
- **World Development Indicators Data:** This [publicly available](#) data is downloaded from the World Bank’s World Development Indicators website. Data version is based on the last update on January 28, 2025. File is named:
 - “*WDI_Raw data.xlsx*”.
- **GDP Per Capita:** This [publicly available](#) data is downloaded from the World Bank’s World Development Indicators website. Data was accessed on November 9, 2025. File is named:
 - “*API_NY.GDP.PCAP.CD_DS2_en_csv_v2_134819.csv*”.
- **Economy Level Enterprise Surveys Data:** This is economy-level aggregated data downloaded from the World Bank’s publicly available [Enterprise Surveys’ website](#). Using the custom query function, this data set was downloaded after selecting all economies and “Gender” as a topic. File is downloaded on November 6, 2025, and named:
 - “*CustomQuery-Nov-6-2025.xlsx*”.
- **Firm Level Enterprise Surveys Data:** This is firm-level full-survey data downloaded from the World Bank’s publicly available [Enterprise Surveys’ data portal](#) on October 28, 2025. Data can be accessed after selecting the page of “Access microdata”. Please see the file name as:
 - “*New_Comprehensive_October_6_2025.dta*”.

Statement about Rights

☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Reproduction Instructions

The materials included in the reproducibility package allow users to reproduce all exhibits. A third party can run the code. Users can reproduce the results by following these steps:

1. Setting the username and the global paths in Part 2 of “0_Master.do”
2. Run main do-file sequentially, and switch from 0 to 1 to run the following do-files in Part 3 of “0_Master.do”:
 - a. preparing external data to be used:
 - i. "1_Data Prep_FLFP.do"
 - ii. "1_Data Prep_gdppc.do"
 - iii. "1_Data Prep_own.do"
 - b. cleaning data sets to be analyzed
 - i. "2_Data Clean_Economy2023.do"
 - ii. "2_Data Clean_EconomyPanel.do"
 - iii. "2_Data Clean_Firm.do"
 - c. create figures and tables
 - i. "3_Data Analysis_Figure 1.do"
 - ii. "3_Data Analysis_Figure 2_Map.do"
 - iii. "3_Data Analysis_Figure 3.do"
 - iv. "3_Data Analysis_Figure 4.do"
 - v. "3_Data Analysis_Table 1.do"
 - vi. "3_Data Analysis_Table 2.do"
 - vii. "3_Data Analysis_Table 3.do"
 - viii. "3_Data Analysis_Annex 1.do"

Directories of Do-Files and List of Exhibits

Do-File Name	Purpose	Output Name, if any	Output Folder
0_Master	Master do-file	/	/
1_Data Prep_FLFP.do	Clean non-agricultural FLFP	WDI_FLFP.dta, WDI_FLFP_2023.dta	"\intermediary"
1_Data Prep_gdppc	Clean gdppc	gdppc.dta, gdppc_2023.dta	"\intermediary"
1_Data Prep_own	Clean ES data on economy-level percent of firms with women ownership participation	ES Female Ownership_2019-23.dta, ES Female Ownership_2019- 23_MostRecent.dta	"\intermediary"
2_Data Clean_Economy2023.do	Create cross-sectional data set for economy- level analysis using WBL 2024 2.0 data, together with other data	WBL2024_forAnalysis.dta	"\output\dtfiles"

2_Data Clean_EconomyPanel.do	Create panel data for economy-level analysis using WBL 1.0 data, together with other data	WBL1- 0_Panel_Cleaned.dta	"\output\dtfiles"
2_Data Clean_Firm.do	Create firm-level panel data set using ES and WBL 1.0 panel data, together with other data	Firm_Level_WBL1- 0_Panel.dta	"\output\dtfiles"
3_Data Analysis_Figure 1.do	Create Figure 1	Figure 1.png	"\output\graphs"
3_Data Analysis_Figure 2_Map.do	Create Figure 2 (Map)	Figure 2.png	"\output\graphs"
3_Data Analysis_Figure 3.do	Create Figure 3	Figure 3.png	"\output\graphs"
3_Data Analysis_Figure 4.do	Create Figure 4	Figure 4.png	"\output\graphs"
3_Data Analysis_Table 1.do	Create Table 1	table1_panel.rtf	"\output\tables"
3_Data Analysis_Table 2.do	Create Table 2	table2_panel.rtf	"\output\tables"
3_Data Analysis_Table 3.do	a) Create Table 3; b) Calculate numbers cited in the section of "Stylized Fact 3"; c) Calculated numbers cited in Annex 4: Summary Statistics	Table 3_logit_regression.doc, Table 3_logit_regression.txt	"\output\tables"
3_Data Analysis_Annex 1.do	Create Annex 1	Annex 1_A.png, Annex 1_B.png	"\output\graphs"

Requirements

Software Requirement:

- Stata version 18
 - spmap
 - distinct

Folder Structure

- **Reproducibility_Entrepreneurship/datawork/dofiles:** This folder includes all the Stata do-files. Details of the do-files are further illustrated in the section of ***“Directories of Do-Files”***
- **Reproducibility_Entrepreneurship/datawork/input:** This folder includes the raw datasets used in the paper. More details are the data sources please see the section of ***“Data Availability Statement/Data Sources”***.
- **Reproducibility_Entrepreneurship/datawork/intermediary:** This folder includes intermediary datasets for use in data cleaning. After the coding scripts are run, this folder should include the following list of data files:
 - ***“EconomyList_WBL_2024_ToUse.dta”***: 190 economies and their associated regions used by the *Women, Business and the Law* (WBL) for the WBL 2024 report. This is extracted from the WBL 2024 Legal Frameworks Score datafile (***“WBL2024_LEGAL_EconLevel_ToPublish.dta”***).
 - ***“ES Female Ownership_2019-23.dta”***: The World Bank’s Enterprise Surveys data on female ownership participation in firms from 2019 to 2023.
 - ***“ES Female Ownership_2019-23_MostRecent.dta”***: Most recently available Enterprise Surveys data on female ownership participation in firms within the years ranging 2019 to 2023.
 - ***“gdppc.dta”***: GDP per capita data across years downloaded from the World Bank’s World Development Indicators.
 - ***“gdppc_2023.dta”***: GDP per capita data, taken the most recently available within the time range of 2022 to 2023.
 - ***“map_WBL_SF_Scores.dta”***: This is a dta file prepared for drawing maps. It includes world boundaries together with WBL’s Supportive Frameworks (SF) score from 2024.
 - ***“WDI_FLFP.dta”***: This is a dta file cleaned on estimating female labor force participation rates for nonagricultural sectors. It also includes data on the female population ages 15-64.
 - ***“WDI_FLFP_2023.dta”***: This is a dta file cleaned on estimating female labor force participation rates for nonagricultural sectors and prepared for cross-sectional analysis on the calendar year of 2023. It also includes data on the female population ages 15-64.
 - ***“world_coords_mercator.dta”***: This is a dta file prepared for creating the map.
 - ***“world_shp_merged.dta”***: This is dta file prepared for creating the map.

- **Reproducibility_Entrepreneurship/datawork/output:** Here includes three sub-folders.
 - **/dtafiles:** This includes the cleaned datasets ready for analysis.
 - “*WBL2024_forAnalysis.dta*”: for conducting cross-sectional analysis using WBL 2024 data, merged with relevant data sets.
 - “*WBL1-0_Panel_Cleaned.dta*”: for conducting panel analysis using WBL 1.0 Panel data, merged with relevant data sets.
 - “*Firm_Level_WBL1-0_Panel.dta*”: for conducting firm-level analysis, merged with WBL 1.0 Panel data.
 - **/graphs:** This includes graphs of Figure 1, 2, 3, and 4, and Annex 1a and 1b.
 - **/tables:** This includes outputs of Table 1, 2, and 3.