

Guidelines for replicating the results in  
**“Does Informal Competition Reduce Firm-Provided Worker Training among Formal Manufacturing SMEs in Sub-Saharan Africa?”**

This draft: May 7, 2026

The guidelines below replicate the results contained in the “P96\_InfCompt-Training\_12-1-2025 JR\_Revised\_Clean.docx” manuscript included in this package.

**Step 1:** Create a folder on your computer where the data, do files, and results will be saved. Call this folder “Training”.

**Step 2:** Obtain the following data from the World Bank’s Enterprise Analysis Unit. Please email Rose Gachina at [rgachina@worldbank.org](mailto:rgachina@worldbank.org) for the data. The datasets you need to ask for are the following:

“New\_Comprehensive\_March\_12\_2024.dta”

“ES-Indicators-Database-Global-Methodology\_March\_12\_2024.dta”

These are primary datasets collected and compiled by the World Bank’s Enterprise Analysis Unit. They are archived and so are not available directly on the website. Please contact me if you have difficulty obtaining these data. The last replication of the results used these data obtained from the Enterprise Surveys team on May 18, 2026.

**Step 3:** Save the two datasets obtained in Step 2 in the folder created in Step 1. Please make sure that the names of these two datasets are as indicated in Step 2.

**Step 4:** You will see a folder called “Files-to-copy” in the replication package. Copy and save this entire folder in the folder you created in Step 1. This sub-folder includes the Stata do files, R-script, and the manuscript.

**Step 5:** Open the folder you created in Step 1 and then open the “Files-to-copy” sub-folder that was created in Step 4. Please check that you have the following files in this sub-folder:

- (i) Basic\_Data-construction.do
- (ii) DGE\_p\_Data-construction.do

- (iii) fc-pc-uc\_Asif.do
- (iv) P96\_Data-Preparation\_5-5-2026.do
- (v) P96\_InfCompt-Training\_12-1-2025 JR\_Revised\_Clean.docx
- (vi) P96\_Inf-Training\_5-5-2026\_Replication.do
- (vii) R-Script\_Placebo-Test1.R

Please do not save any other file in this sub-folder.

**Step 6:** You need to download raw data from World Bank's Informal Economy Database, which is used for replicating Figure 1. Instructions on what to do if you are unable to download these data from the original source are provided below. For now, follow these steps to obtain the raw data from the original source:

- (i) Go to this website: <https://www.worldbank.org/en/research/brief/informal-economy-database>
- (ii) On the right side of your computer screen and just below “Downloads”, you will see a link called “[Informal economy data \(Excel\)](#)”. Click on this link. It will open an excel file.
- (iii) In this excel file, go the “DGE\_p” worksheet by clicking on this worksheet. You will see the top row as Economy, Code, 1990, 1991, ..., 2020.
- (iv) Manually change the top row by replacing 1990, 1991, ..., 2020 in the top row with v\_1990, v\_1991, ..., v\_2020 (that is, just add “v\_” before 1990-2020 in the top row). This ensures that Stata reads the top row as variable names instead of numbers or data.
- (v) Now save the edited excel file as “DGE\_Raw.xlsx” in the folder created in Step 1. Please make sure that this is saved as “xlsx” and not “xls” file. You may close the excel file now.

If you downloaded the data successfully, please proceed to the next step. If you were unable to download the data for some reason, the data are provided in the replication package. Find the file “DGE\_Raw.xlsx” in the “Raw-data” folder included in the replication package. Now copy and paste the file into the folder you created in Step 1. This file was downloaded on 5/5/2026.

**Step 7:** Now, you need to download data for the regions/continents.

- (i) For this, please go this website: <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>
- (vi) In the upper half of your computer screen, find and click on the link “[current classification by income in XLSX format](#)”. This will download an excel file on your computer. Open this excel file and save it as “Regions\_Raw.xlsx” in the folder created in Step 1. You may now close this excel file.

If you downloaded the data successfully, please proceed to the next step. If you were unable to download the data for some reason, the data are provided in the replication package. Find the file

“Regions\_Raw.xlsx” in the “Raw-data” folder included in the replication package. Now copy and paste the file into the folder you created in Step 1. This file was downloaded on 5/5/2026.

**Step 8:** Cleaning-up the raw data obtained in steps 6 and 7.

- (i) Open the do file “DGE\_p\_Data-construction.do” that you saved in Step 5 above. On line 4 of the do file, replace “Folder path” with the full directory/path of the Training folder that you created in Step 1. Save the file now.
- (vii) Run this do file from start to finish. This will generate the DGE\_p.dta file that will be used to replicate Figure 1. It will also generate “Regions-wbcode-WDI.dta” file that will be used later to add regions to the countries. Both these data files will be automatically saved in the folder created in Step 1.
- (viii) Close the do file. Clear and exit Stata.

**Step 9:** Construct the Basic\_for\_March\_12\_2024-Comp\_3-28-2024.dta data file

In this step, you will compile data that contains basic information about Enterprise Surveys, such as the year of survey, latest round or earlier round, official country name, and so on. For this, follow these steps:

- (i) Open the do file “Basic\_Data-construction.do” that you saved in Step 5.
- (ii) On line 4 of the do file, replace “Folder path” with the full directory/path of the Training folder created in Step 1. Save the file now.
- (iii) Run the do file from start to finish. It should take less than 1 minute to run.
- (iv) After the do file finishes running, close the do file. Clear and exit Stata.

**Step 10:** In this step, you will download raw data from World Development Indicators and Doing Business project. If you are unable to download the raw data for some reason, these are provided in the replication package as indicated below. For now, follow these steps to obtain the raw data from the source:

- (i) Go the Data Development Platform (DDP) website:  
<https://databank.worldbank.org/source/world-development-indicators>
- (ii) On the left side of the screen, click on “Database” and select “World Development Indicators”.
- (iii) Just below “Database”, find and click on “Country”. In the menu that opens, select all countries. You can do so by a single click on the box with a yes tick mark (just below Country). You should see “Selected 266” (or close to 266) on your screen in the row for “Country”.
- (iv) Just below, “County”, click on “Series”. Unselect any previously selected series by clicking on the box with a small cross just below “Series”. You should now see “Selected 0” in the row for “Series”.

- (v) Now, from the drop-down menu, select the variables/series with the following **exact** match:
  - GDP per capita, PPP (constant 2021 international \$)
  - GDP per capita growth (annual %)
  - Population ages 15-64, female
  - Population ages 00-04, female (% of female population)
  - Population ages 00-04, male (% of male population)
  - Population ages 05-09, female (% of female population)
  - Population ages 05-09, male (% of male population)
  - Population, female
  - Population, male
  - Population, total
  - School enrollment, primary (% gross)

Many series have similar names. So, please be sure that your selected series match exactly with the ones listed above.

- (vi) You should see “Selected 11” in the “Series” row now.
- (vii) Just below “Series”, find and click on “Time” and select all the years available. You can do so by a single click on the box with a yes tick mark (just below “Time”).
- (viii) On the right side of the screen, click on “Apply Changes”.
- (ix) On the top right side, click on “Download options” and “Advanced options” from the drop-down menu.
- (x) In the box that opens on your screen, select “.xlsx” for Excel format, “List” for the Data format, and “Blank” for the NA preference. Leave the rest to default settings.
- (xi) Click on “Download” in the box. This will download an excel file on your computer. Open and save this excel file as “WDI.xlsx” in the folder you created in Step 1.
- (xii) Close the excel file. Close the browser.

If you successfully downloaded the data, you may proceed to Step 11. If the data did not download properly for some reason, then skip step 10 completely and instead copy and save the file “WDI.xlsx” file from the “Raw-data” folder included in the replication package to the folder you created in Step 1. These data were downloaded on 5/5/2026.

### **Step 11: Download the Doing Business data**

In this step, you will repeat Step 10 but download a different series. For this, follow these steps:

- (i) Go the Data Development Platform (DDP) website:  
<https://databank.worldbank.org/source/world-development-indicators>
- (ii) On the left side of the screen, click on “Database” and select “Doing Business”.

- (iii) Just below “Database”, find and click on “Country”. In the menu that opens, select all countries. You can do so by a single click on the box with a yes tick mark (just below Country). This should show “Selected 192” (or close to 192) on your screen in the row for “Country”.
- (iv) Just below, “Country”, click on “Series”. Unselect any previously selected series by clicking on the box with a small cross just below “Series”. You should now see “Selected 0” in the row of “Series”.
- (v) Now, from the drop-down menu, select the variable/series with this **exact** match: “Starting a business: Cost – Men (% of income per capita)”
- (vi) You should now see “Selected 1” in the “Series” row.
- (vii) Just below “Series”, find and click on “Time” and select all the years available. You can do so by a single click on the box with a yes tick mark (just below “Time”).
- (viii) On the right side of the screen, click on “Apply Changes”.
- (ix) On the top right side, click on “Download options” and “Advanced options” from the drop-down menu.
- (x) In the box that opens on your screen, select “.xlsx” for Excel format, “List” for the Data format, and “Blank” for the NA preference. Leave the rest to default settings.
- (xi) Click on “Download” in the box. This will download an excel file on your computer. Open and save this excel file as “DoingBusiness.xlsx” in the folder you created in Step 1.
- (xii) Close the excel file. Close the browser.

If you successfully downloaded the data in this step, you may proceed to Step 12. If the data did not download properly for some reason, then skip step 11 completely and instead copy and save the file “DoingBusiness.xlsx” from the “Raw-data” folder provided to the folder you created in Step 1. These data were downloaded on 5/5/2026.

## Step 12: Download Rule of Law data

In this step, you will repeat Step 10 but download a different series. For this, do the following:

- (i) Go the Data Development Platform (DDP) website:  
<https://databank.worldbank.org/source/world-development-indicators>
- (ii) On the left side of the screen, click on “Database” and select “Worldwide Governance Indicators”.
- (iii) Just below “Database”, find and click on “Country”. In the menu that opens, select all countries. You can do so by a single click on the box with a yes tick mark (just below Country). This should show “Selected 216” (or close to 216) on your screen in the row for “Country”.

- (iv) Just below, “Country”, click on “Series”. Unselect any previously selected series by clicking on the box with a small cross just below “Series”. You should now see “Selected 0” in the row of “Series”.
- (v) Now, from the drop-down menu, select the variable/series with this **exact** match: “Rule of Law – Governance estimate (approx. -2.5 to +2.5)”
- (vi) You should now see “Selected 1” in the “Series” row.
- (vii) Just below “Series”, find and click on “Time” and select all the years available. You can do so by a single click on the box with a yes tick mark (just below “Time”).
- (viii) On the right side of the screen, click on “Apply Changes”.
- (ix) On the top right side, click on “Download options” and “Advanced options” from the drop-down menu.
- (x) In the box that opens on your screen, select “.xlsx” for Excel format, “List” for the Data format, and “Blank” for the NA preference. Leave the rest to default settings.
- (xi) Click on “Download” in the box. This will download an excel file on your computer. Open and save this excel file as “Rule-of-Law.xlsx” in the folder you created in Step 1.
- (xii) Close the excel file. Close the browser.

If you successfully downloaded the data, you may proceed to Step 13. If the data did not download properly, skip Step 12 completely and instead copy and paste the file “Rule-of-Law.xlsx” from the “Raw-data” folder provided to the folder you created in Step 1. These data were downloaded on 5/5/2026.

### **Step 13: Download the Economic Freedom of the World Data**

In this step, you will download the Economic Freedom of the World data. Follow these steps:

- (i) Go this website: <https://efotw.org/economic-freedom/dataset?geozone=world&min-year=2&max-year=0&filter=0&page=dataset>
- (ii) On the top right side of the screen, click on “Download Entire Dataset”. This will download an excel file on your computer. Open and save this file as “EFW.xlsx” in the folder created in Step 1. Note that the file may be in “protected view”, in which case click on “Enable Editing” at the top of the excel file before saving it.
- (iii) After saving the excel file, close the file and the browser.

If you successfully downloaded the data in this step, you may proceed to Step 14. If the data did not download properly, skip step 13 completely and instead copy and paste the file “EFW.xlsx” from the “Raw-data” folder provided to the folder you created in Step 1. These data were downloaded on 5/5/2026.

All data downloads are now complete.

#### Step 14: Data cleaning

- (i) Open the do file “P96\_Data-Preparation\_5-5-2026.do”, which you saved in Step 5.
- (ii) On line 7 of the do file, replace “Folder path” with the full directory/path of the Training folder that you created in Step 1. Save the do file now.
- (iii) Run the do file from start to finish. It should take about 1 minute to run. It will create a dataset that will be used in the regression analysis. If needed, you can run it multiple times. Earlier versions of the files created by this do file will be replaced by the ones created by the last run.
- (iv) Close the do file. Exit Stata.

#### Step 15: Data for Figure 5

- (i) Open R-Studio.
- (ii) You need to have “haven” and “dplyr” packages installed in your R-Studio. Install these packages if you do not have them. These can be installed by running the following codes from the R-Studio console:  
`install.packages("dplyr")`  
`install.packages("haven")`

The versions of the above packages used are: package ‘dplyr’ was built under R version 4.5.3 and package ‘haven’ was built under R version 4.5.3.

- (iii) Now, open the R-script “R-Script\_Placebo-Test1.R” which you saved in Step 5.
- (iv) On line 10 of the script, replace “Folder path” with the full directory/path of the folder you created in Step 1. Please note that you should use forward slashes (/) instead of the usual backward slashes in specifying the folder directory.
- (v) Run this R script from start to finish.
- (vi) If needed, you can run the file multiple times. Earlier versions of the files created by this R-script will be replaced by ones created by the last one.
- (vii) Save the file and close it. Exit R-Studio.

#### Step 16: Stata programs

Please check that the following programs are installed in your Stata. For any missing programs, you can download them by running “ssc install xx” from the Stata window (after replacing xx with the program name below):

“cmp” (version 8.7.9 18 April 2024)  
“moremata”  
“kdens” (version 2.0.2 04aug2007 Ben Jann)  
“xml\_tab” (version 3.35 17Oct2007 M. Lokshin, Z. Sajaia)  
“avplot” (version 3.5.0 17jun2009)  
“marginsplot” (version 1.3.5 17oct2023}  
“fracreg” (version 1.1.2 02mar2022}  
“putexcel” (version 2.2.7 13feb2024}  
“ivreg2” (version 4.1.12 14aug2024)  
“putdocx” (version version 1.9.0 15sep2025)

**Step 17:** Create a sub-folder called “Results” within the folder that you created in Step 1. Replication results from the next step will be save in this “Results” folder.

### **Step 18: Results replication**

- (i) Open the “P96\_Inf-Training\_5-5-2026\_Replication.do” file in Stata. You had saved this file in Step 5.
- (ii) On line 5 of the do file, replace “Folder path” with the full directory/path of the Training folder that you created in Step 1. Save the do file now.
- (iii) Run this do file from start to finish in one go. If it stops in between, run the full file again. It should take about 13 minutes to run all the codes. Once all the codes are run, you will see “You may exit now” at the bottom of your screen.
- (iv) Close the do file. Clear and exit Stata. All the estimations are done now.

You can run this do file as many times as needed. Results from the previous runs will be replaced by results created from the last one.

All the figures and tables that you need to match are now saved in the “Results” sub-folder.

### **Step 19: Match results**

Open the “Results” sub-folder that you created in Step 17. Here you will see all the figures and tables. You can match these with the corresponding ones in the manuscript. The manuscript is the file “P96\_InfCompt-Training\_12-1-2025 JR\_Revised\_Clean.docx”, which is included in the

replication package. You had saved this file in Step 5. Follow these instructions to properly match the figures and tables in the manuscript with the ones created by the above instructions.

### Figures:

For the figures, you will see the following “.png” files in the “Results” sub-folder: Figure 1, Figure 2a, Figure 2b, Figure 2c, Figure 2d, Figure 3a, Figure 3b, Figure 4a, Figure 4b, and Figure 5. Match these figures with the corresponding ones in the manuscript. In the manuscript, all figures are placed just after the References.

### Tables:

Most tables are saved in the excel file called “Informal-train.xls”, and one table each is saved in “Table 3\_F-stat.xlsx”, “Table 4\_F-stat\_Hansen.xlsx”, “Table\_A1\_Final.docx”, and “Table\_A3\_Summary\_Statistics.docx”.

Many tables were manually formatted somewhat. For example:

- (i) the tables in the manuscript have “Yes” entries for the fixed effects, but you will not see these in the replicated tables. These “yes” entries were manually inserted
- (ii) you will see some “(dropped)” entries in the replicated tables that were manually deleted before inserting the table in the manuscript
- (iii) the order of the rows in the manuscript may be slightly different from the ones in the replicated tables. This is because in some tables, some rows were manually moved up or down for presentation purposes.

Nevertheless, you will be able to match all the numbers in the manuscript tables with the ones in the replicated tables. Please make sure that you compare the same variable (listed in the first column in each table) in manuscript and the replicated tables.

You can match the results in manuscript tables with the ones replicated by the steps above as follows:

| Table in the manuscript                | Matched table in the “Results” sub-folder                                                                       |
|----------------------------------------|-----------------------------------------------------------------------------------------------------------------|
| Table 1                                | See worksheet “table 1” in “Informal-train.xls” file in the Results sub-folder.                                 |
| Table 2                                | See worksheet “table 2” in “Informal-train.xls” file in the Results sub-folder.                                 |
| Table 3, Panel A                       | See worksheet “table 3_Panel A” in “Informal-train.xls” file in the Results sub-folder.                         |
| Table 3, Panel B (except the last row) | See worksheet “table 3_Panel B” in “Informal-train.xls” file in the Results sub-folder..                        |
| Table 3, last row for “F-statistics”   | See “Kleibergen-Paap rk Wald F statistic” values in “Table 3_F-stat.xlsx” file saved in the Results sub-folder. |

|                                               |                                                                                                               |
|-----------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| Table 4, Panel A                              | See worksheet “table 4_Panel A” in “Informal-train.xls” file in the Results sub-folder.                       |
| Table 4, Panel B (except the last three rows) | See worksheet “table 4_Panel B” in “Informal-train.xls” file in the Results sub-folder.                       |
| Table 4, last three rows                      | See the first row and the last two rows in “Table 4_F-stat_Hansen.xlsx” file saved in the Results sub-folder. |
| Table 5                                       | See worksheet “table 5” in “Informal-train.xls” file in the Results sub-folder.                               |
| Table 6                                       | See worksheet “table 6” in “Informal-train.xls” file in the Results sub-folder.                               |
| Table A1                                      | See “Table_A1_Final.docx” file saved in the Results sub-folder.                                               |
| Table A3                                      | See “Table_A3_Summary_Statistics.docx” file saved in the Results sub-folder.                                  |
| Table A4                                      | See worksheet “table A4” in “Informal-train.xls” file saved in the Results sub-folder.                        |
| Table A5                                      | See worksheet “table A5” in “Informal-train.xls” file saved in the Results sub-folder.                        |
| Table A6                                      | See worksheet “table A6” in “Informal-train.xls” file saved in the Results sub-folder.                        |
| Table A7                                      | See worksheet “table A7” in “Informal-train.xls” file saved in the Results sub-folder.                        |
| Table A8                                      | See worksheet “table A8” in “Informal-train.xls” file saved in the Results sub-folder.                        |
| Table A9                                      | See worksheet “table A9” in “Informal-train.xls” file saved in the Results sub-folder.                        |

Stata version used: StataNow/MP 19.5 for Windows (64-bit x86-64), Revision 28 Jan 2026

R-Studio version used: 2026.1.2.418

**Please feel free to contact me at [mamin@worldbank.org](mailto:mamin@worldbank.org) if you have any questions.**

**Thank you.**