

## Read me file: LEVERAGING TRADE FOR MORE AND BETTER JOBS

**Instructions:** Open “ReplicationMasterReport\_GenDatav1”, change the project directory and run the code. Once it finishes, open “ReplicationMasterReportv4”, change the project directory and run the code.

**/Data** contains the raw files of Trade in Employment (TiM) 2021, Trade in Value Added (TiVA) 2021, TiM 2023, TiVA 2023 that are used to create two main datasets for different years. It also contains the WDI raw files and Other data files. The final datasets used for the remaining analysis in the cited papers are also included, namely Activities in Exports (AiE), Routine Task Intensity (RTI), and Trade & Gender, and can be downloaded from the replication packages shared by the authors. The information for each raw data file is the following:

- “.../AiE”: Activities in exports datasets, from the replication package of the background paper [“Export Diversification from an Activity Perspective: An Exploration Using Occupation Data”](#).
  - Available for download [here](#).
  - Availability: Public and included in the package.
- “.../Other Data”: contains other data to merge with the main datasets:
  - “/Cntrysample2021”: Initial country sample for TiM and TiVA 2021.
    - Available for download [here](#): Go to documentation, then “Appendix A-TiM Country list”.
    - Availability: Public.
    - Downloaded on: June 2, 2024.
  - “/country\_codes\_piaac”: Countries’ iso3, iso2 and full name for piaac available countries.
    - Available for download [here](#).
    - Availability: Public.
    - Downloaded on: March 17, 2024.
  - “/countrynames”: Countries’ full name.
    - Available for download [here](#).
    - Availability: Public.
    - Downloaded on: December 6, 2023.
  - “/dist\_cepii”: Border information across countries.
    - Available for download [here](#). Go to download, click on stata and open the “Gravity” file, keep variable “contig” and data only from 1995.
    - Availability: Public and included in the package.
    - Downloaded on: November 30, 2018.
  - “/GGDC\_extra”: Additional countries with information on jobs in exports.
    - Availability: Shared by the authors of “Export Diversification from an Activity Perspective: An Exploration Using Occupation Data”, included in the package.
  - “/Income”: Countries income classification by year.
    - Available for download [here](#). Go to “historical classification by income in XLSX format.”, sheet “Country Analytical History”.
    - Availability: Public.
    - Downloaded on: August 31, 2022.
  - “/Indcode\_new\_old”: Correspondence table between industry codes.
    - Available for download [here](#). Go to “TiM old versus new codes”, last table.
    - Availability: Public
    - Downloaded on: December 6, 2023.
  - “/ISIC4\_ETD\_EORA”: Correspondence table among industry codes.
    - Built by the team based on:
      - [ETD-ISIC Rev 4concordance](#)
      - ISIC-EORA concordance: Built by the team, as no official ISIC-EORA concordance is available (see documentation [here](#)).
    - Availability: Public and included in the package.
    - Downloaded on: October 4, 2022.

- “/maritime\_frontiers\_islands”: Maritime frontiers across countries.
  - Manually created based on google maps.
  - Availability: Public and included in the package.
  - Downloaded on: January 19, 2023.
- “/Taxonomy\_new\_only\_final”: GVC taxonomy group by year based on WDR 2020.
  - Available for download [here](#). Go to Chapter 1, download data, then go to “Map 1.1” sheet in “WDR2020\_chapter1\_data” file.
  - Availability: Public.
  - Downloaded on: August 31, 2022.
- “.../RTI”: Routine Task Intensity dataset from the background paper [“The Role of Global Value Chains for Worker Tasks and Wage Inequality”](#). This dataset is created from the replication package of [“Technology, Skills, and Globalization: Explaining International Differences in Routine and Nonroutine Work Using Survey Data”](#). The following steps describe the process to create the final dataset based on the replication package mentioned above.
  - Follow the replication package:
    - Download [PIAAC](#), and [STEP](#) data (the replication package also mentions CULS which is not needed for this analysis), the country specific source is specified in their replication package. Cyprus data is not available anymore, so it is not considered in this analysis.
    - Combine PIAAC data: Run the code “1a. Prepare combined PIAAC update”, which creates “combined\_piaac.dta”.
    - Combine STEP data: Run the code “1b. Prepare combined STEP”, which creates “combined\_step.dta”.
    - Join both datasets: Run the code “2. Join PIAAC, STEP and other data”, which creates “joint\_sample.dta”.
    - Create “piaac\_and\_step\_employed\_fdilit.dta” dataset: Run “predata\_a1\_datasource.do” from the auxiliary do-files by running line 108 from “4. Analysis.do”.
    - Create RTI variables: Run “predata\_a2\_clean.do” from the auxiliary do-files by running line 110 from “predata\_a2\_clean.do”. In lines 230-243 RTI is calculated.
    - Prepare “dataset.dta”: Go to “Code” and Run “0\_EORA\_piaac” and “1\_eora\_to\_piaac\_”. Set “data\_input” and “data\_output” to “Data/RTI”. Using “piaac\_and\_step\_employed\_fdilit.dta” (define sectors and occupation groups, merge EORA data, keep relevant variables).
  - Available for download: [Replication package](#).
  - Availability: For any questions regarding this dataset please contact Piotr Lewandowski [piotr.lewandowski@ibs.org.pl](mailto:piotr.lewandowski@ibs.org.pl).
- “.../TiM2021/Raw files”: Individual csv files by variable.
  - Available for download [here](#) and [here](#).
  - Availability: Public.
  - Downloaded on: June 3, 2024.
- “.../TiM2023/Raw files”: Individual csv files by variable.
  - Available for download [here](#) and [here](#).
  - Availability: Public.
  - Downloaded on: December 8, 2023.
- “.../TiVA2021/Raw files”: Individual csv files by variable.
  - Available for download [here](#).
  - Availability: Public.
  - Downloaded on: June 3, 2024.
- “.../TiVA2023/Raw files”: Individual csv files by variable.
  - Available for download [here](#) and [here](#).
  - Availability: Public.

- Downloaded on: November 21, 2023.
- “.../Trade & Gender”:
  - Available for download: [here](#).
    - Firm-Level-TFP-Estimates-and-Factor-Ratios-Data-and-Documentation
    - StandarizedNew2006-2018-core4
  - Availability: Public and included in the package.
  - Downloaded on: April 9, 2018.
- “.../WDIdata”:
  - Available for download [here](#).
  - Availability: Public.
  - Downloaded on: January 7, 2022.

**/Code** contains 3 do files:

- “ReplicationMasteReportGenData” generates the main datasets and other datasets that start from the main one. Two main datasets are created from raw files for TIM and TiVA using different data releases (2021 and 2023). It also generates other datasets that transform and merge WDI data.
- “sector\_cntry\_deflators” creates the country and sector deflators used to compute real variables at different levels of disaggregation. It generates two different sets of deflators, one set for the 2021 data release and another set for the 2023 data release. It is called at the start of “ReplicationMasterReport\_GenData”.
- “ReplicationMasterReport” creates Figures 1, Figure 6, Figure 9, and Figure 10. It also creates the excel files needed to create the figures produced in excel (Figures 2, Figure 3, Figure 4, Figure 5, Figure 7 and Figure 8).

**/Excel plots:** Contains Figure 2, Figure 3, Figure 4, Figure 5, Figure 7, and Figure 8. It uses the data in the excel files creates in the folders with the name of the corresponding figure. Data in the Figure folder has to be copied into the FigureX\_plots to produce the plots.

**/Appendix:** Contains the full regression tables for Figure 3, Figure 4, Figure 5, and Figure 7.