

Beneath the Surface: Deep Trade Agreements and the Potential of a Deeper AfCFTA

LIONEL FONTAGNÉ* ANTHONIN LEVELU†
JEAN-CHRISTOPHE MAUR‡ NADIA ROCHA§ GIANLUCA SANTONI¶

July 7, 2026

README : Replication package.

*Paris School of Economics, 48 boulevard Jourdan, 75014 Paris; E-mail: lionel.fontagne@psemail.eu – Lionel Fontagné thanks the support of the EUR grant ANR-17-EURE-0001 and the i-MIP institute.

†Joint Research Center (European Commission), Via E. Fermi, 2749, 21027 Ispra; Italy; E-mail: anthonin.levelu@ec.europa.eu

‡The World Bank, 1818 H St NW, Washington, DC 20433, United States; E-mail: jmaur@worldbank.org

§Inter American Development Bank, 1300 New York Avenue, NW Washington, DC 20577, United States; E-mail: nadiaro@iadb.org

¶Paris School of Economics, 48 boulevard Jourdan, 75014 Paris; E-mail: gianluca.santoni@psemail.eu – Gianluca Santoni thanks the support of the EUR grant ANR-17-EURE-0001 and the i-MIP institute.

Disclaimer: This paper is part of an ongoing World Bank project and is therefore subject to modification. This research was supported by the World Bank Umbrella Facility for Trade trust fund, which receives contributions from the governments of the Netherlands, Norway, Sweden, Switzerland and the United Kingdom. The perspectives and opinions presented in this paper are solely those of the authors and do not necessarily represent the views of their respective institutions.

Summary

This replication package reproduces the results reported in the paper. This 'readme' file details the structure of the package and the steps required to run it. Scripts in different programming languages can be run individually, although preferably in the same order. Please note that minor discrepancies may arise when running the code across different operating systems, particularly for Table 2.

Data availability statement & provenance statements

- This paper involves analysis of external data.
- I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- All data are publicly available, but also provided directly in this package.

Computational requirements

The package was last ran on MacOS Monterey v12.6, M2 chip (8-core CPU), 2022, 16GB RAM. All R and Stata needed packages are automatically installed by the scripts.

Software used

- Python 3.7, using a Python native virtual environment
- Stata SE 17
- R (release RStudio-2022.07.2-576)

Runtime and storage

- Approximate time needed to reproduce the analysis: 8-9 hours
- Approximate storage space needed: 9 GB - 30 GB

Acknowledgements

This package was prepared by the corresponding author. Its preparation has benefited from the authors' support. Corresponding author: anthonin.levelu@ec.europa.eu

Program files (folder /0.prog)

This folder contains all scripts, running all steps needed to produce the results of the paper. Languages used are Python, Stata, and R.

- **FLMRS_replication.py:**

This is the main Python script. Users should follow the steps as indicated in the file, install the needed Python packages, and modify the object 'root' (i.e., the path where this replication package is stored) as well as the Stata and R instance locations with their actual paths before running the script. The coding structure is in 7 main steps to improve clarity.

- 0- It runs the Stata scripts for some Stata settings.
- 1- It runs the Stata scripts for data cleaning and preparation of the Deep PTA data.
- 2- It runs the Stata scripts for data cleaning and preparation of the Trade data.
- 3- It runs the Shapley value analysis in Python.
- 4- It runs the Multi-Objective Counterfactual methodology in R.
- 5- It runs the partial equilibrium analysis results.
- 6- It runs the results of the GE analysis.
- 7- It runs descriptive statistics results and miscellaneous results.

- **00_Stata_settings.do:**

Stata script for Stata settings.

- **./1_DTA:**

Contains the Stata scripts to run the DTA database preparation and perform the clustering analysis.

- **./2_Trade:**

Contains the Stata scripts to prepare the Trade data.

- **./3_Shapley:**

Contains the Python script to run the Shapley analysis.

- **./4_MOC:**

Contains the R script to run the MOC analysis.

- **5_pe_analysis.do:**

Stata script to run the partial equilibrium analysis.

- **./6_ge:**

Contains the Stata scripts to run the GEPPML analysis.

- **7_misc.do:**

Stata script to run additional results, such as descriptive statistics tables.

Data files (folder **/1.data**)

This folder contains all the needed datasets.

- **./class:**

Contains HS classification concordance between millesime.

- **./COMTRADE:**

Contains data from COMTRADE.

- **./CTY:**

Contains some country identifiers and additional information.

- **./DTA:**

Contains the Deep Trade Agreement database from the World Bank, notably the vertical and horizontal versions.

- **./FAO:**

Contains agriculture trade data from FAO.

- **./GRAVITY:**

Contains Gravity variables sourced from CEPII.

- **./LANG:**

Contains shared language variable sourced from CEPII.

- **./OTHERS:**

Contains some files not elsewhere stored, notably the lists the agreement retained in the analysis, their predicted cluster, and whether these involve African economies and data from the World Bank's DTA database vertical.

- **./STATAMAP:**

Stata files to create world maps.

- **./TARIFFS:**

Contains tariffs data sourced from WITS and TRADEPROD.

- **./temp:**

This is a folder storing temporary data.

- **./trade_costs:**

Contains some trade costs data, notably the transport cost from the OECD.

- **./TradeProd_CEPII:**

Contains production data from TRADEPROD.

- **TradeProd_Data_Toolkit_2024.dta:**

This is the main .dta file used in most of the regressions.

- **./UNIDO:**

Contains manufacturing trade data from UNIDO.

Output files (folder **/2.res**)

This folder contains graphs and tables of results produced by the different scripts in multiple formats. It is organized as in the /0.prog folder.

Data sources and reference

- **./class:**

UNSD (2023), "Classifications on economic statistics: All HS codes and descriptions". Retrieved from: <https://unstats.un.org/unsd/classifications/Econ>

- **./COMTRADE:**

Replication of Mayer et al. (2023) combining import (CIF value minus maritime transport costs) and export (FOB) declarations. Import and export data for 1962-2022 for all reporters and all partners (SITC Rev 1 classification) was sourced from COMTRADE retrieved from <https://comtrade.un.org/>. It was then converted to ISIC classification (using SITC Rev 1 to ISIC Rev 3.1 concordance sourced from WITS https://wits.worldbank.org/product_concordance.html and aggregated for industry groupings. The maritime transport costs come from the OECD's Maritime Transport Cost Database available at <https://stats.oecd.org/Index.aspx?DataSetCode=MTC>.

- **./CTY:**

Data on country identifiers from the World Bank, and FAO.

- **./DTA:**

Obtained from the World Bank Deep Trade Agreements Database at <https://datatopics.worldbank.org/dta/table.html>. Source: Mattoo et al. (2020)

- **./FAO:**

Obtained from WITS, retrieved from: https://wits.worldbank.org/product_concordance.html

- **./GRAVITY:**

Obtained from CEPII gravity database at https://www.cepii.fr/ceprii/en/bdd_modele/bdd_modele_item.asp?id=8. Source: Conte et al. (2022)

- **./LANG:**

Obtained from CEPII gravity database at https://www.cepii.fr/ceprii/en/bdd_modele/bdd_modele_item.asp?id=8. Source: Conte et al. (2022).

- **./OTHERS:**

Own elaboration:

- The file 'Cluster_list_2024.xlsx' lists the agreements retained in the analysis, their predicted cluster, and whether these involve African economies. To reproduce the file, one can start from "p3_compute_cluster_DTA.do", line 457.

- The file 'id_agree_bilat_all.dta' is derived from ". /DTA/DTA 2.0 - Vertical Content (v2).xlsx", sheet 'Bilateral Information' from the World Bank Deep Trade Agreement Database.

- The file 'country_classification.xlsx' is retrieved from The World Bank, World Development Indicators: <https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-a.html>

- **./STATAMAP:**

Stata maps. Retrieved from: https://www.stathelp.se/en/spmap_world_en.html.

- **./TARIFFS:**

Obtained from TRADEPROD. Retrieved from: https://www.cepii.fr/cepii/en/bdd_modele/bdd_modele_item.asp?id=5

- **./trade_costs:**

Organisation for Economic Co-operation and Development. (2024). International transport and insurance costs of merchandise trade (ITIC). OECD Data. <https://www.oecd.org/en/data/datasets/international-transport-and-insurance-costs-of-merchandise-trade-itic.html>

- **./TradeProd_CEPH:**

The CEPH trade and production database. Source: Mayer et al. (2023). Retrieved from: https://www.cepii.fr/cepii/en/bdd_modele/bdd_modele_item.asp?id=5

- **./UNIDO:**

Obtained from UNIDO, "Annual manufacturing trade database". Retrieved from: <https://stat.unido.org/data/table?dataset=mtd#data-browser>.

Bibliography

- Conte, M., P. Cotterlaz, and T. Mayer (2022, July). The cepii gravity database. Working Papers 2022-05, CEPII.
- Mattoo, A., N. Rocha, and M. Ruta (2020). *Handbook of deep trade agreements*. World Bank Publications.
- Mayer, T., G. Santoni, and V. Vicard (2023). *The CEPII trade and production database*. CEPII.