

Replication Package for “What’s in a Name? Implications of Extensive Margin Measurement in International Trade”

Overview

This package includes all the replication materials for “What’s in a Name? Implications of Extensive Margin Measurement in International Trade” by Ana Fernandes, Alejandro Forero, Devaki Ghose, and Piyush Panigrahi

Data Availability Statement

The authors of the manuscript certify they have legitimate access to and permission to use the data used in this manuscript. Data files stored in folders `data/raw/trade` and `data/raw/allidscorrespondence` are proprietary and are not available for public sharing. Proprietary shipment-level export and import data were obtained from the S&P Panjiva trade data platform for seven countries: India (March 2024), Colombia (August 2024), Mexico (July and August 2024), Peru (January 2024), Sri Lanka (August 2024), Uruguay (December 2023), and Vietnam (November 2023).

Repository Structure

The repository is organized in the following folders: `root` | `├── data/` | `├── raw/` # raw data files | `├── processed/` # intermediate files generated during the analysis | `└── output/` # final data outputs from the analysis | `├── code/` # scripts for data cleaning, analysis and result generation | `├── output/` | `├── figs/` # figures generated for the paper | `└── tabs/` # tables generated for the paper | `└── README.md`

Folder Organization

The contents of the folders are as follows:

1. **Raw Data (data/raw)**

Contains the original unprocessed datasets.

- `trade/` (contains firm-level trade datasets)
 - Source: S&P Panjiva
 - License: Proprietary
- `support/` (contains auxilliary datasets)
 - `indicator_names.csv`
 - Source: Own Conventions
 - License: Public
 - `consolidation_1996_2022.dta`
 - Source: Exporter Dynamics Database 3.0 (forthcoming):
<https://datacatalog.worldbank.org/search/dataset/0042326/Exporter-Dynamics-Database>

- License: Public
 - Contact: Ana M. Fernandes (afernandes@worldbank.org)
- statistics_exporterimporters_2019.xlsx
 - Source: Exporter Dynamics Database 3.0 (forthcoming)" <https://datacatalog.worldbank.org/search/dataset/0042326/Exporter-Dynamics-Database>
 - License: Public
 - Contact: Ana M. Fernandes (afernandes@worldbank.org)
- allidscorrespondence/ (contains the mapping between cleaned and platform IDs)
 - Source: S&P Panjiva and own algorithms
 - License: Proprietary

2. **Processed Data (data/processed)**

Holds intermediate datasets created during data cleaning and transformation steps.

- formatted/ (contains formatted datasets; not for public disclosure)
- dynamics/ (contains augmented datasets; not for public disclosure)
- firm_outcomes/ (contains firm-level outcomes datasets; not for public disclosure)
- indicators/
- indicator_differences/
- id_mapping/ (contains firm names and identifiers; not for public disclosure)
- trade_share_matching/
- misidentified_trade.csv
- relship_tenure_E2F.csv
- relship_tenure_I2F.csv
- indicator_diff_multiproduct_E.csv
- indicator_diff_agedist_E.csv
- indicator_diff_agedist_I.csv

3. **Output Data (data/output)**

Stores the final indicator datasets produced by the analysis and ready for use in reporting. The column names in these files are

Indicator,Country,Value_Cleaned,Value_Platform,Year.

- indicators_E.csv (firm-level exports indicators)
- indicators_I.csv (firm-level imports indicators)
- indicators_E2F.csv (firm-to-firm exports indicators)
- indicators_I2F.csv (firm-to-firm imports indicators)

4. **Code (code/)**

Includes scripts for data preprocessing, analysis, and result generation.

- !master.R (master code to run all analysis)
- format_data/ (formats raw data into a common format; cannot be executed using disclosed data)

- !master.R
- format_data.R
- format_data_USA.R
- get_varnames.R
- load_ctry_names.R
- create_dynamics_data/ (augments formatted data with dynamics information; cannot be executed using disclosed data)
 - !master.R
 - create_dynamics_data_E.R
 - create_dynamics_data_I.R
 - create_dynamics_data_E2F.R
 - create_dynamics_data_I2F.R
 - create_dynamics_data_EP.R
 - create_dynamics_data_IP.R
 - create_dynamics_data_E2FP.R
 - create_dynamics_data_I2FP.R
 - create_dynamics_data_ED.R
 - create_dynamics_data_IO.R
 - create_dynamics_data_USA.R
- compute_firm_outcomes/ (computes firm-level outcomes; cannot be executed using disclosed data)
 - !master.R
 - compute_firm_outcomes_E.R (computes firm-level exports outcomes)
 - compute_firm_outcomes_I.R (computes firm-level imports outcomes)
 - compute_firm_outcomes_E2F.R (computes firm-to-firm exports outcomes)
 - compute_firm_outcomes_I2F.R (computes firm-to-firm imports outcomes)
- compute_indicators/ (computes aggregate indicators; cannot be executed using disclosed data)
 - !master.R
 - compute_indicators_E.R (computes indicators for firm-level exports)
 - compute_indicators_I.R (computes indicators for firm-level imports)
 - compute_indicators_E2F.R (computes indicators for firm-to-firm exports)
 - compute_indicators_I2F.R (computes indicators for firm-to-firm exports)
 - export_indicators_output.R

- eval_misreported_trade/ (analysis for Table 3; cannot be executed using disclosed data)
 - !master.R
 - analyze_id_mapping.R
 - compute_misreported_trade.R
- compute_indicator_diff.R (analysis for Tables 4, 6, 13)
- compute_multiproduct_exports.R (analysis for Table 5; cannot be executed using disclosed data)
- compute_linktradeshare.R (analysis for Tables 11, 12; cannot be executed using disclosed data)
- compare_distributions/ (analysis for Tables A1–A4; cannot be executed using disclosed data)
 - !master.R
 - compare_distributions_E.R
 - compare_distributions_I.R
 - compare_distributions_E2F.R
 - compare_distributions_I2F.R
- compute_age_dist.R (analysis for Tables A5, A6; cannot be executed using disclosed data)
- compute_tenure_dist.R (analysis for Tables A7, A8; cannot be executed using disclosed data)
- plot_bar_indicators/
 - !master.R
 - plot_bar_indicators_E.R (creates Figures 1–4)
 - plot_bar_indicators_I.R (creates Figures A1–A3)
 - plot_bar_indicators_E2F.R (creates Figure A4)
 - plot_bar_indicators_I2F.R (creates Figure A4)
- make_tables/
 - !master.R
 - tab_data_coverage.R (generates Table 1)
 - tab_misreported_trade.R (generates Table 3)
 - tab_indicator_diff.R (generates Tables 4, 6, 14)
 - tab_multiproduct_exports.R (generates Table 5)
 - tab_compute_xmpremia_diff.R (generates Tables 7, 8; cannot be executed using disclosed data)
 - tab_compare_ground_truth.R (generates Tables 9, 10)
 - tab_matching_types.R (generates Tables 11, 12)
 - tab_distribution_diff.R (generates Tables A1–A4)
 - tab_agedist.R (generates Tables A5, A6)
 - tab_tenure_dist.R (generates Tables A7, A8)

5. **Figures (output/figs)**

Contains plots used in the paper.

- Figure 1:
 - indicator_comp_E_A1.pdf
 - indicator_comp_E_A6i.pdf
 - indicator_comp_E_A1pc.pdf
 - indicator_comp_E_A6iii.pdf
- Figure 2:
 - indicator_comp_E_B1.pdf
 - indicator_comp_E_B7ii.pdf
- Figure 3:
 - indicator_comp_E_C1.pdf
 - indicator_comp_E_C2.pdf
 - indicator_comp_E_C4.pdf
- Figure 4:
 - indicator_comp_E_B3i.pdf
- Figure A1:
 - indicator_comp_I_A1.pdf
 - indicator_comp_I_A6i.pdf
 - indicator_comp_I_A1pc.pdf
 - indicator_comp_I_A6iii.pdf
- Figure A2:
 - indicator_comp_I_B1.pdf
 - indicator_comp_I_B7ii.pdf
- Figure A3:
 - indicator_comp_I_C1.pdf
 - indicator_comp_I_C2.pdf
 - indicator_comp_I_C4.pdf
- Figure A4:
 - indicator_comp_E2F_A1.pdf
 - indicator_comp_E2F_C1.pdf
 - indicator_comp_E2F_B8i.pdf
 - indicator_comp_I2F_A1.pdf
 - indicator_comp_I2F_C1.pdf
 - indicator_comp_I2F_B8i.pdf

6. **Tables (output/tabs)**

Contains LaTeX table outputs used in the paper.

- Table 1: data_coverage.tex
- Table 3: groundtruth_indicator_comp_E_select.tex
- Table 4: groundtruth_indicator_comp_I_select.tex

- Table 5: `misreported_trade_by_type.tex`
- Table 6: `indicator_diff_E_select.tex`
- Table 7: `indicator_diff_multiproduct_E.tex`
- Table 8: `indicator_diff_I_select.tex`
- Table 9: `xmtrade_aggregate.tex`
- Table 10: `xmpremia_E_select.tex`
- Table 11: `matching_type_E2F.tex`
- Table 12: `matching_type_I2F.tex`
- Table 13:
 - `indicator_diff_E2F_select.tex`
 - `indicator_diff_I2F_select.tex`
- Table A1: `distribution_diff_E_select.tex`
- Table A2: `distribution_diff_I_select.tex`
- Table A3: `distribution_diff_E2F_select.tex`
- Table A4: `distribution_diff_I2F_select.tex`
- Table A5: `indicator_diff_agedist_E.tex`
- Table A6: `indicator_diff_agedist_I.tex`
- Table A7: `indicator_diff_relshiptenure_E2F.tex`
- Table A8: `indicator_diff_relshiptenure_I2F.tex`

Usage

To replicate the results, run the master script `code/!master.R` from the root folder as the working directory. This master script will call the necessary sub-scripts in sequence.