



What's in a Name? Implications of Extensive Margin Mismeasurement in International Trade

Fourth Submission: RR_WLD_2025_452

Ceren Huryol, Maria Reyes Retana

reproducibility@worldbank.org

January 27, 2026

This review verifies the reproducibility of the exhibits included in the paper “*What's in a Name? Implications of Extensive Margin Mismeasurement in International Trade*”.

Contents in this review:

1. Main findings
2. List of exhibits and reproducibility status
3. Reproduction Environment

Main findings

- **Every exhibit in the main paper and 10 randomly selected exhibits in the appendix were reproduced accurately**
- The code was successfully executed on a new computer after:
 1. Opening the master script and running the code.
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code takes approximately 10 hours to run.
- We conducted our reproducibility analysis based on the paper shared by the authors on January 14, 2026.
- **Verification Process and Data Handling:**
 - The reproducibility package relies on 2 types of data: open and restricted data.
 - Open data is included in the public reproducibility package. Data is forthcoming in the Development Data Hub.
 - Reviewers used restricted data provided directly by the authors to conduct the reproducibility verification, and this is not included in the public reproducibility package.
 - *data_hash_report.csv* lists the SHA256 hashes of all files in the Data folder to support data integrity checks. Users who acquire the forthcoming/accessible/limited-access/restricted data can use this file to verify that the data has not been altered.
- **Reproducibility Summary:**

- **Data:** Some data is restricted and has not been included in the reproducibility package. For more details, please refer to the README file.
- **Code:** All code files are included in the reproducibility package but operate from intermediate data, not raw data.
- **Outputs:** All outputs are generated by code included in the reproducibility package.
- **Dependencies environment:** The reviewers created a new environment using the latest versions of dependencies available at the moment of the review.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Figure 1 Reproduced.** Compared against the following files in output/figs:
 - indicator_comp_E_A1.pdf
 - indicator_comp_E_A6i.pdf
 - indicator_comp_E_A1pc.pdf
 - indicator_comp_E_A6iii.pdf
- **Figure 2 Reproduced.** Compared against the following files in output/figs:
 - indicator_comp_E_B1.pdf
 - indicator_comp_E_B7ii.pdf
- **Figure 3 Reproduced.** Compared against the following files in output/figs:
 - indicator_comp_E_C1.pdf
 - indicator_comp_E_C2.pdf
 - indicator_comp_E_C4.pdf
- **Figure 4 Reproduced.** Compared against output/figs/indicator_comp_E_B3i.pdf.
- **Table 1 Reproduced.** Compared against *data_coverage.tex*.
- **Table 2** Does not show analysis results.
- **Table 3 Reproduced.** Compared against *groundtruth_indicator_comp_E_select.tex* in output/tabs.
- **Table 4 Reproduced.** Compared against *groundtruth_indicator_comp_I_select.tex* in output/tabs.
- **Table 5 Reproduced.** Compared against *misreported_trade_by_type.tex* in output/tabs.
- **Table 6 Reproduced.** Compared against *indicator_diff_E_select.tex* in output/tabs.
- **Table 7 Reproduced.** Compared against *indicator_diff_multiproduct_E.tex* in output/tabs.
- **Table 8 Reproduced.** Compared against *indicator_diff_I_select.tex* in output/tabs.
- **Table 9 Reproduced.** Compared against *xmtrade_aggregate.tex* in output/tabs.

- **Table 10 Reproduced.** Compared against *xmpremia_E_select.tex* in output/tabs.
- **Table 11 Reproduced.** Compared against *matching_type_E2F.tex* in output/tabs.
- **Table 12 Reproduced.** Compared against *matching_type_I2F.tex* in output/tabs.
- **Table 13 Reproduced.** Compared against *indicator_diff_E2F_select.tex* and *indicator_diff_I2F_select.tex* in output/tabs.

Results in the Annex

For the Appendix, we did not review every exhibit. Instead, we randomly selected 10 exhibits from the remaining datasets to assess the appendix. Our review was based on those 10 exhibits. Since they were chosen randomly, we operate under the assumption that if all randomly selected exhibits are reproducible, then the rest should be as well. The seed used to generate the random selection was 8479045, the Stata version used was 18, and the exhibits selected were: 1, 3, 4, 5, 7, 8, 9, 10, 11, and 12.

Code used for selecting exhibits:

https://github.com/worldbank/wb-reproducible-research-repository/blob/main/resources/select_appendix_exhibits.do

- **Table A1 Reproduced.** Compared against *distribution_diff_E_select.tex* in output/tabs.
- **Table A3 Results reproduced, but table includes manual changes from the code output.** Compared against *distribution_diff_E2F_select.tex* in output/tabs.
- **Table A4 Reproduced.** Compared against *distribution_diff_I2F_select.tex* in output/tabs.
- **Table A5 Reproduced.** Compared against *indicator_diff_agedist_E.tex* in output/tabs.
- **Table A7 Reproduced.** Compared against *indicator_diff_relshiptenure_E2F.tex* in output/tabs.
- **Table A8 Reproduced.** Compared against *indicator_diff_relshiptenure_I2F.tex* in output/tabs.
- **Figure A1 Reproduced.** Compared against the following files in output/figs:
 - *indicator_comp_I_A1.pdf*
 - *indicator_comp_I_A1pc.pdf*
 - *indicator_comp_I_A6iii.pdf*
 - *indicator_comp_I_A6i.pdf*
- **Figure A2 Reproduced.** Compared against the following files in output/figs:
 - *indicator_comp_I_B1.pdf*
 - *indicator_comp_I_B7ii.pdf*
- **Figure A3 Reproduced.** Compared against the following files in output/figs:
 - *indicator_comp_I_C1.pdf*
 - *indicator_comp_I_C2.pdf*

- indicator_comp_I_C4.pdf
- **Figure A4 Reproduced.** Compared against the following files in output/figs:
 - 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

Reproduction Environment

Paper exhibits were reproduced on a computer with the following specifications:

- OS: Windows 11 Enterprise
- Processor: Intel(R) Core(TM) i5-1145G7 CPU @ 2.60GHz
- Memory available: 15.7 GB
- Software version: R 4.4.1