



“Bridging the Gap in Trade Reporting: Insights from the Discrepancy Index”

First Submission: RR_WLD_2024_111

Yukiko Suzuki

reproducibility@worldbank.org

May 13th, 2024

This review verifies the reproducibility of the exhibits included in the paper *“Bridging the Gap in Trade Reporting: Insights from the Discrepancy Index”*.

Contents in this review:

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

Main findings

- The code was successfully executed on a databricks environment.
- This paper relies on raw data that is constantly updated. Additionally, we did not have access to the exact snapshot of the raw data used to produce the exhibits in the paper. For these reasons, we utilized a snapshot of an intermediate dataset (an indicator dataset constructed by the authors) to verify the reproducibility of the exhibits.
- We also verified that the indicator dataset constructed from the most recent raw dataset is comparable to the intermediate dataset snapshot provided by the authors.
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code takes approximately 1 hours to run.
- We conducted our reproducibility analysis based on the paper shared by the authors on github in May.
- Every exhibit has been reproduced accurately.
- **Reproducibility Summary:**
 - **Data:** All data sources are publicly available but not all are included in the package.
 - **Code:** All code files (from cleaning to analysis) are included in the package
 - **Outputs:** All outputs are generated by code included in the reproducibility package.
 - **Reproducibility verification:** Not all of the data used to conduct this verification is included in the public package.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Figure 1** Does not apply.
- **Figure 2** Reproduced.
- **Figure 3** Reproduced.
- **Figure 4** Reproduced.
- **Figure 5** Reproduced.
- **Figure 6** Reproduced.
- **Figure 7** Reproduced.
- **Figure 8** Reproduced.
- **Table 1** Does not apply.
- **Figure 9** Reproduced.
- **Table 2** Reproduced.
- **Figure 10** Reproduced.
- **Figure 11** Reproduced.
- **Figure 12** Reproduced.
- **Figure 13** Reproduced.
- **Figure 14** Reproduced.

Results in the Annex

- **Table 3** Reproduced.
- **Table 4** Reproduced.
- **Table 5** Reproduced.
- **Table 6** Reproduced.
- **Table 7** Reproduced.
- **Table 8** Reproduced.
- **Table 9** Reproduced.
- **Table 10** Reproduced.
- **Table 11** Reproduced.

- Table 12 Reproduced.
- Table 13 Reproduced.
- Table 14 Reproduced.
- Figure 15 Reproduced.
- Figure 16 Reproduced.
- Figure 17 Reproduced.
- Figure 18 Reproduced.
- Figure 19 Reproduced.
- Figure 20 Reproduced.
- Figure 21 Does not apply.
- Figure 22 Reproduced.
- Figure 23 Reproduced.
- Figure 24 Reproduced.
- Figure 25 Reproduced.
- Figure 26 Reproduced.
- Figure 27 Reproduced.
- Figure 28 Reproduced.
- Figure 29 Reproduced.
- Figure 30 Reproduced.
- Figure 31 Reproduced.
- Figure 32 Reproduced.
- Figure 33 Reproduced.
- Table 15 Reproduced.
- Table 16 Reproduced.
- Table 17 Does not apply.
- Table 18 Does not apply.
- Table 19 Does not apply.
- Table 20 Does not apply.

Reproduction Environment

- Paper exhibits were reproduced in a computer with the following specifications:
 - Databricks Runtime Version: 14.3 LTS (includes Apache Spark 3.5.0, Scala 2.12)
 - Worker Type: Standard_E4ds_v4 (32 GB Memory, 4 cores)
 - Python version: 3.10.12