



Innovation Patterns in the World Bank Group Portfolio

First Submission: RR_WLD_2026_639

Ankriti Singh

reproducibility@worldbank.org

May 5, 2026

This review verifies the reproducibility of the exhibits included in the paper "*Innovation Patterns in the World Bank Group Portfolio*".

Contents in this review:

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

Main findings

- **Every exhibit of the paper has been reproduced accurately.**
- The code was successfully executed on a new computer after:
 1. Restoring the environment in "requirements.txt"
 2. Running "scripts/run_analysis.py"
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code takes approximately 1 minute to run.
- We conducted our reproducibility analysis based on the paper shared by the authors in the package on May 1, 2026.
- **Verification Process and Data Handling:**
 - The reproducibility package relies on 1 types of data: open.
 - Open data is included in the public reproducibility package.
 - The reproducibility verification used the author-constructed keyword dictionaries as inputs, without independently replicating the LLM-assisted steps used to produce them. The methods and prompt records for these steps are documented by the authors in the methodology/ folder. Verification covered the fully automated Python pipeline, executed by running scripts/run_analysis.py.
 - *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:** All data sources are publicly available and included in the reproducibility package.
- **Code:** All code files (from cleaning to analysis) are included in the reproducibility package.
- **Outputs:** All outputs are generated by code included in the reproducibility package.
- **Dependencies environment:** The reviewers reproduced an existing environment for dependencies using dependency files or an environment metadata file provided by the authors.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- Table 3.1 Reproduced.
- Table 3.2 Reproduced.
- Table 3.3 Reproduced.
- Table 3.4 Does not apply. Not an analytical output.
- Table 3.5 Reproduced.
- Table 3.6 Reproduced.
- Table 3.7 Reproduced.
- Table 3.8 Reproduced.
- Table 3.9 Reproduced.
- Table 3.11 Reproduced.
- Table 3.12 Reproduced.
- Table 3.13 Reproduced.
- Table 3.14 Does not apply. Not an analytical output.
- Table 3.15 Reproduced.
- Table 3.16 Reproduced.
- Figure 3.1 Reproduced.
- Figure 3.2 Reproduced.
- Figure 3.3 Does not apply. Not an analytical output.
- Figure 3.4 Reproduced.
- Table 4.1 Reproduced.
- Table 4.2 Does not apply. Not an analytical output.

- **Figure 4.1** Reproduced.
- **Table 5.1** Does not apply. Not an analytical output.

Results in the Annex

- **Table D.1** Reproduced.
- **Table D.2** Reproduced. Includes parts a-d.
- **Figure D.1** Reproduced.
- **Table E.1** Reproduced.
- **Table E.2** Reproduced.
- **Table E.5** Reproduced.
- **Table E.6** Does not apply. Not an analytical output.
- **Table E.10** Reproduced.
- **Table E.11** Does not apply. Not an analytical output.
- **Table F.1** Reproduced.
- **Table F.2** Reproduced.
- **Table F.3** Reproduced.
- **Table F.4** Reproduced.
- **Table F.5** Reproduced.

Reproduction Environment

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

- OS: Windows 11 Enterprise
- Processor: INTEL(R) XEON(R) PLATINUM 8562Y+ (2.80 GHz) (4 processors)
- Memory available: 32.0 GB
- Software version: Python 3.13.2