



Reproducible Research Repository

Fusing data sources to measure multidimensional poverty

First Submission: RR_WLD_2026_600

Innocent Bushayija Sano

reproducibility@worldbank.org

May 28, 2026

This review verifies the reproducibility of the exhibits included in the paper “*Fusing data sources to measure multidimensional poverty*”.

Contents in this review:

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

Main findings

- **Every exhibit in the main section of the paper and 10 randomly selected exhibits in the appendix were reproduced accurately**
- The code was successfully executed on a new computer after:
 1. Doing changing the root directory, and installing dependencies.
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code takes approximately 13 hours to run.
- We conducted our reproducibility analysis based on the paper shared by the authors by email on May 21st paper.
- **Verification Process and Data Handling:**
 - The reproducibility package relies on 1 type of data: restricted data (available only from the data owners).
 - Reviewers used limited-access 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:** All data is limited-access and has not been included in the reproducibility package. For more details, refer to the README file.

- **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 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

- **Table 1** Does not apply. Table 1 is a non-analytical output.
- **Table 2** Does not apply. Table 2 is a non-analytical output.
- **Table 3** Does not apply. Table 3 is a non-analytical output.
- **Table 4** Reproduced. Table 4 contained a numerical discrepancy: the MPIW value under Corrected (global) for the pce-rews scenario in the MPM panel shows 6.09 in the paper but 6.32 in the code output. The numerical difference is due to a platform dependency from the “highs” package used for linear programming solver (optimization) which uses platform-specific linear algebra libraries (BLAS/LAPACK).
- **Figure 1** Does not apply. Figure 1 is a non-analytical output.
- **Figure 2** Does not apply. Figure 2 is a non-analytical output.
- **Figure 3** Does not apply. Figure 3 is a non-analytical output.
- **Figure 4** Does not apply. Figure 4 is a non-analytical output.
- **Figure 5** Does not apply. Figure 5 is a non-analytical output.
- **Figure 6** Does not apply. Figure 6 is a non-analytical output.
- **Figure 7** Does not apply. Figure 7 is a non-analytical output.
- **Figure 8** Reproduced.
- **Figure 9** Reproduced.
- **Figure 10** Reproduced.
- **Figure 11** Reproduced.
- **Figure 12** Reproduced.
- **Figure 13** Reproduced.

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 are operating 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 645466, the Stata version used was 19.0 MP, and the exhibits selected were: Table A1, Table A2, Figure A2, Figure A3, Figure A5, Figure A6, Figure A7, Figure A8, Table A5, Table A6.

- **Table A1** Reproduced.
- **Table A2** Reproduced.
- **Figure A2** Reproduced.
- **Figure A3** Reproduced.
- **Figure A5** Reproduced.
- **Figure A6** Reproduced.
- **Figure A7** Reproduced.
- **Figure A8** Reproduced.
- **Table A5** Reproduced. Table A5 contains slight numerical differences: the MAPE value under Corrected (local) for the e-ws scenario at the National level shows 84.23 in the paper but 84.26 in the code output (= 0.03). However, since this is below the 1.0 percent threshold, this is considered reproducible.
- **Table A6** Reproduced. Table A6 contains slight numerical differences: the MAPE value under Corrected (local) for the e-ws scenario at the National level shows 64.15 in the paper but 64.17 in the code output (= 0.02); and the MAPE value under Corrected (local) for the ew-es scenario at the Sub-national level shows 18.06 in the paper but 18.08 in the code output (= 0.02). However, since this is below the 1.0 percent threshold, this is considered reproducible.

Reproduction Environment

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

- OS: Windows 11 Enterprise
- Processor: Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz (2.90 GHz) (2 processors)
- Memory available: 16.0 GB
- Software version: R 4.5.2