



Global Poverty Estimation Using Private and Public Sector Big Data Sources

First Submission: PP_WLD_2024_67-v01

Maria Reyes Retana

reproducibility@worldbank.org

February 28th, 2024

This review verifies the reproducibility of the exhibits included in the paper “*Global Poverty Estimation Using Private and Public Sector Big Data Sources*”.

Contents in this review:

1. Main findings
2. Reproducibility assessment
3. List of exhibits and reproducibility status

Main findings

- The code was successfully executed on a new computer after:
 1. Installing the needed packages.
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code takes approximately 2 hours to run.
- Every exhibit has been reproduced accurately.

Reproducibility assessment

- Paper exhibits were attempted to be reproduced in 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 4.3.0
- **Reproducibility Summary:**
 - **Data:** Some data is confidential and has not been included in the package. For more details, please refer to the README file.
 - **Code:** The package includes all necessary code files, which primarily operate on intermediate data. Although the code required to process the raw data is provided, our verification is based on the intermediate data.

- **Outputs:** All outputs are generated by the code included in the reproducibility package.
- **Reproducibility verification:** The review process used data provided directly by the authors to conduct the reproducibility verification, and this is not included in the package.
- We conducted our reproducibility analysis based on the paper shared by the authors Dropbox folder. Our validation involved comparing the results generated by the code with the exhibits in the paper to assess their consistency.
- This verification process began with an intermediate dataset provided by the authors. The authors include the necessary code to construct this intermediate dataset within the reproducibility package. Due to the constraints related to data access and the substantial time required to execute the data construction code—approximately five months for the Facebook data and two weeks for other sources—the verification of the code for constructing the intermediate dataset in terms of computational reproducibility was not feasible. However, it was confirmed that the package contains a detailed and complete description of all raw data sources, including the requirements for access.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Table 1** Does not apply.
- **Figure 1 Reproduces.** Figure: cor_all_fb_osm.png
- **Figure 2 Reproduces.** Figure: cor_changes.png
- **Figure 3 Reproduces.** Figure: performance_country_avg_types.png
- **Figure 4 Reproduces.** Figure: global_scatter_map.png
- **Figure 5 Reproduces.** Figure: explain_var_in_r2.png
- **Figure 6 Reproduces.** Figure: changes_results.png
- **Figure 7 Reproduces.** Figure: panel a ml_changes_explain.png
- **Figure 8 Reproduces.** Figure: policy_exp_nigeria_gadm_2.png
- **Figure 9 Reproduces.** Figure: example_daytime_images.png

Results in the Annex

- **Table S1 Reproduces.** Figure: dhs_summary_table.tex
- **Table S2 Reproduces.** Figure: dhs_summary_table_changes.tex
- **Figure S1 Reproduces.** Figure: dhsindex_globalindex_cor.png
- **Figure S2 Reproduces.** Figure: fb_features_cor_each_country.png
- **Figure S3 Reproduces.** Figure: cor_between_vars.png
- **Figure S4 Reproduces.** Figure: policy_exp_nigeria_gadm_2.png
- **Figure S5 Reproduces.** Figure: r2_scatter_eachcountry_districtagg.png
- **Figure S6 Reproduces.** Figure: r2_pooled_by_continent.png
- **Figure S7 Reproduces.** Figure: country_featureset_r2.png
- **Figure S8 Reproduces.** Figure: changes_scatter_cluster.png
- **Figure S9 Reproduces.** Figure: changes_scatter_adm2.png
- **Table S3 Reproduces.** Figure: explain_error_levels_lm.tex
- **Table S4 Reproduces.** Figure: explain_error_changes_lm.tex
- **Figure S10 Reproduces.** Figure: explain_error_levels.png
- **Figure S11 Reproduces.** Figure: explain_error_changes.png
- **Figure S12 Reproduces.** Figure: policy_exp_nigeria_gadm_1.png
- **Figure S13 Reproduces.** Figure: ml_results_type.png
- **Table S5 Reproduces.** Figure: within_across_pca_sd.tex
- **Table S6 Does not apply.**
- **Figure S14 Reproduces.** Figure: lsms_pov_measure_cor.png
- **Figure S15 Reproduces.** Figure: lsms_r2.png
- **Figure S16 Reproduces.** Figure: lsms_feature_type.png
- **Figure S17 Reproduces.** Figure: educ_fb_dhs_boxplot.png
- **Figure S18 Reproduces.** Figure: educ_fb_dhs_scatter_cluster.png
- **Figure S19 Reproduces.** Figure: educ_fb_dhs_scatter_adm2.png
- **Figure S20 Reproduces.** Figure: educ_fb_dhs_explain.png