Reproducible Research Repository

Urban Job Accessibility for the Poor

First Submission: FR_WLD_2026_690
Ankriti Singh, Zulekha Shoib
reproducibility@worldbank.org
August 13, 2026

This review verifies the reproducibility of the exhibits included in the paper *"Urban Job Accessibility for the Poor"*.

Contents in this review:

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 partially reproduced on a new computer. Part 1 was successfully executed. Part 2 analysis scripts were successfully executed. The pre-computed workflows for South Africa, Philippines, Part 2 feature preparation, and Hanoi were not executed by the replicators, as these depend on routing engines and system libraries that are not fully portable. See below for details.
- The code was successfully executed on a new computer after:
  1. Opening the R project "Job_Accessibility_Index.Rproj"
  2. Opening the 00.master.R and running the code.
- The following steps were **not** executed by the replicators:
  - Pre-computed workflows for South Africa, Philippines, Part 2 data preparation, and Hanoi - deliberately not re-run as their outputs depend on OSRM and r5r routing engines and system GDAL/PROJ libraries that are not pinned by `renv` and vary across machines.
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code takes approximately 60 minutes to run.
- We conducted our reproducibility analysis based on the paper shared by the authors on July 30th, 2026.

- **Verification Process and Data Handling**:
  - The reproducibility package relies on 4 types of data: open, accessible (forthcoming to World Bank Development Data Hub), limited-access (available after registering), and restricted data (available only from the data owners).
  - Open data is included in the public reproducibility package.
  - Reviewers used forthcoming/accessible/restricted data provided directly by the authors to conduct the reproducibility verification, and this is not included in the public reproducibility package. Some of the data is forthcoming to World Bank Development Data Hub.
  - The pre-computed workflows for South Africa, Philippines, Part 2 feature preparation, and Hanoi are included in the master script as commented-out code and were not executed by the replicators. These workflows produce travel-time matrices using OSRM and `r5r`, and spatial features using `sf` and its GDAL/PROJ dependencies. Results from these steps vary with the routing engine build, the underlying OpenStreetMap extract, and system library versions, none of which are pinned by `renv`. The authors froze these outputs and treated the upstream scripts as documentation. The replicators verified that the frozen outputs are present in the package and are consumed correctly by the executed scripts.
  - *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 (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 1** Does not apply. The exhibit is an not an analytical output.
- **Figure 1** Reproduced.
- **Figure 2** Reproduced.
- **Figure 3** Reproduced.
- **Figure 4** Reproduced.
- **Figure 5** Reproduced.
- **Figure 6** Reproduced.

- **Figure 7** Reproduced.
- **Figure 8** Reproduced.
- **Table 2** Reproduced.
- **Figure 9** Reproduced.
- **Figure 10** Reproduced.
- **Figure 11** Reproduced.
- **Figure 12** 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 899522, the Stata version used was 17, and the exhibits selected were: Tables A1, A2, A3, Figures A1, A2, A3, A8, A9, A12, A14.

- **Table A1** Does not apply. The exhibit is an not an analytical output.
- **Table A2** Reproduced.
- **Figure A1** Reproduced.
- **Figure A2** Reproduced.
- **Figure A3** Reproduced.
- **Table A3** Reproduced.
- **Figure A8** Reproduced.
- **Figure A9** Reproduced.
- **Figure A12** Reproduced.
- **Figure A14** 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) (2 processors)
- Memory available: 128.0 GB
- Software version: R 4.5.2.