



Measuring Job Accessibility: Different Methods and New Data

Second Submission: RR_MDG_2025_389

Luis Eduardo San Martin, Shreya Suthakar

reproducibility@worldbank.org

25th July, 2025

This review verifies the reproducibility of the exhibits included in the paper “*Measuring Job Accessibility: Different Methods and New Data*”.

Contents in this review:

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

Main findings

- **Every exhibit has been reproduced accurately.**
- The project has two parts:
 1. Geospatial data processing in ArcGIS, which was verified over a virtual meeting on July 18th, 2025 and an in-person meeting on July 24th, 2025.
 2. A Stata do-file which was executed by the reviewers. The do-file ran after updating the global path with the new folder location.
- The output of the Stata do-file demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The data processing in ArcGIS takes approximately 4 hours to process and the Stata code takes 10 minutes to run.
- We conducted our reproducibility analysis based on the paper shared by the authors via OneDrive on July 09th, 2025.
- **Verification Process and Data Handling:**
 - The reproducibility package includes 3 types of data: open, accessible (public but not included in the reproducibility package), and restricted data (available only to the authors).
 - Open data is included in the public reproducibility package.
 - Reviewers used processed data from all data sources provided directly by the authors to run the Stata part of the reproducibility verification. This data is also included in the public reproducibility package.

- Reviewers verified that the limited-access data matches what was provided by the authors only for the *Measuring Urban Living Standards in Antananarivo 2016* dataset; results of this verification are documented in the file `comparison_report.csv`.
- *data_hash_report.csv* lists the SHA256 hashes of all input files in the Data folder to support data integrity checks. Users who acquire the accessible or 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 are included in the reproducibility package but operate from intermediate data, not raw data. The intermediate data is produced with manual steps for geospatial data processing using ArcGIS. A high-level description of the geospatial data processing is included in the README, along with history logs of the ArcGIS data processing.
 - **Dependencies environment:** The reviewers created a new environment for Stata 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

- **Figure 1 Reproduced.** The exhibit was reproduced by the reviewers using the Excel file *fig_1.xlsx*.
- **Figure 2 Reproduced.** The exhibit was generated manually in ArcGIS during a virtual meeting. Its instructions are included in the README.
- **Figure 3 Reproduced.** The exhibit was reproduced by the reviewers using the Stata do-file *Tana_Access_vs_Firms_2025feb24_OpenData.do*.
- **Figure 4 Reproduced.** The exhibit was reproduced by the reviewers using the Stata do-file *Tana_Access_vs_Firms_2025feb24_OpenData.do*.
- **Figure 5 Reproduced.** The exhibit was reproduced by the reviewers using the Stata do-file *Tana_Access_vs_Firms_2025feb24_OpenData.do*.
- **Figure 6 Reproduced.** The exhibit was reproduced by the reviewers using the Stata do-file *Tana_Access_vs_Firms_2025feb24_OpenData.do*.
- **Figure 7 Reproduced.** The exhibit was reproduced by the reviewers using the Stata do-file *Tana_Access_vs_Firms_2025feb24_OpenData.do*.
- **Figure 8 Reproduced.** The exhibit was generated manually in ArcGIS during a virtual meeting. Its instructions are included in the README.
- **Figure 9 Reproduced.** The exhibit was generated manually in ArcGIS during a virtual meeting. Its instructions are included in the README.

- **Figure 10 Reproduced.** The exhibit was generated manually in ArcGIS during a virtual meeting. Its instructions are included in the README.
- **Figure 11 Reproduced.** The exhibit was generated manually in ArcGIS during a virtual meeting. Its instructions are included in the README.
- **Table 1 Reproduced.** The exhibit was reproduced by the reviewers using the Stata do-file *Tana_Access_vs_Firms_2025feb24_OpenData.do*. The results were compared against lines 198 to 210 of *log.txt*.
- **Table 2 Reproduced.** The exhibit was reproduced by the reviewers using the Stata do-file *Tana_Access_vs_Firms_2025feb24_OpenData.do*. The results were compared against lines 217-224 of *log.txt*.
- **Table 3 Reproduced.** The exhibit was reproduced by the reviewers using the Stata do-file *Tana_Access_vs_Firms_2025feb24_OpenData.do*. The results were compared against lines 803, 915, 1027, 1139, 1251, 1363, 1400-1462 of *log.txt*.
- **Table 4 Reproduced.** The exhibit was reproduced by the reviewers using the Stata do-file *Tana_Access_vs_Firms_2025feb24_OpenData.do*. Column 1 was compared against lines 238-258, 295, and 309 of *log.txt*. Column 2 was compared against lines 318-338, 376, and 389 of *log.txt*. Column 3 was compared against lines 398-418, 456, and 469 of *log.txt*. Column 4 was compared against lines 478-498, 536, and 549 of *log.txt*. Column 5 was compared against lines 558-578, 616, and 629 of *log.txt*. Column 6 was compared against lines 658-658, 696, and 709 of *log.txt*.
- **Table 5 Reproduced.** The exhibit was reproduced by the reviewers using the Stata do-file *Tana_Access_vs_Firms_2025feb24_OpenData.do*. Column 1 was compared against lines 724-745, 783, 808, 821, and 833 of *log.txt*. Column 2 was compared against lines 836-857, 894, 920, 933, and 945 of *log.txt*. Column 3 was compared against lines 948-969, 1007, 1032, 1045, and 1057 of *log.txt*. Column 4 was compared against lines 1060-1081, 1119, 1144, 1157, and 1169 of *log.txt*. Column 5 was compared against lines 1172-1193, 1231, 1256, 1269, and 1281 of *log.txt*. Column 6 was compared against lines 1284-1306, 1343, 1368, 1381, and 1393 of *log.txt*.
- **Table 6 Reproduced.** The exhibit was reproduced by the reviewers using the Stata do-file *Tana_Access_vs_Firms_2025feb24_OpenData.do*. Column 1 was compared against lines 1475-1494, 1531, 1536, 1559, and 2370 of *log.txt*. Column 2 was compared against lines 1570-1589, 1627, 1632, 1655, and 2673 of *log.txt*. Column 3 was compared against lines 1666-1685, 1723, 1728, 1751, and 2976 of *log.txt*. Column 4 was compared against lines 1762-1781, 1819, 1824, 1847, and 3279 of *log.txt*. Column 5 was compared against lines 1858-1877, 1915, 1920, 1943, and 3582 of *log.txt*. Column 6 was compared against 1954-1973, 2011, 2016, 2039, and 3885 of *log.txt*.

Reproduction Environment

Paper exhibits were reproduced using two computers with the following specifications:

- Computer 1 (ArcGIS geospatial data processing):

- OS: Windows 11 Enterprise, version 23H2
 - Processor: Intel(R) Core(TM) i5-114567 @2.60GHz 1.50GHz
 - Memory available: 15.7 GB
 - Software version: ArcGIS Desktop 10.8.1
- Computer 2 (Stata):
 - OS: Windows 11 Enterprise, version 23H2
 - Processor: Intel(R) Xeon(R) Platinum 8562Y+ 2.80 GHz (2 processors)
 - Memory available: 32.0 GB
 - Software version: Stata (StataNow) 18.5 MP-Parallel Edition