



Reproducible Research Repository

Public and Private Transit: Evidence from Lagos

Fourth Submission: RR_NGA_2025_298

Xuerong Lei, Luis Eduardo San Martin

reproducibility@worldbank.org

August 29th, 2025

This review verifies the reproducibility of the exhibits included in the paper “*Public and Private Transit: Evidence from Lagos*”.

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 code was successfully executed on a new computer after:
 1. Changing the working directory in the Stata do-files.
 2. Changing the directory result in the function `getProjectRoot()` of *configurePaths.py*.
 3. Changing the working directory in *11B_checkinShift.R*.
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code takes approximately 20 hours to run.
- We conducted our reproducibility analysis based on the paper shared by the authors using the project's GitHub repository on August 12, 2025.
- **Verification Process and Data Handling:**
 - The reproducibility package relies on 4 types of data: open, accessible (public but not included in the reproducibility package), forthcoming (to be made available in the future), and restricted data (available only from the authors).
 - Open data is included in the public reproducibility package (datasets “*Lagos privately operated transport routes*” and “*Global Human Settlement Layer (GHSL)*”).
 - Reviewers verified that the accessible data (dataset “*COVID-19 Community Mobility Reports*”) matches what was provided by the authors; results of this verification are documented in the file *comparison_report.csv*.

- Forthcoming data will be made available through the World Bank Development Data Hub (dataset “Public transportation trips data for Lagos for June 2020 to December 2023”). Reviewers used data provided directly by the authors to conduct the reproducibility verification, 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 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 (from cleaning to analysis) are included in the reproducibility package.
 - **Outputs:** Some outputs are not generated by code (created manually in QGIS) but instructions for producing these outputs are included in the README file.
 - **Dependencies environment:**
 - * **Stata, R, and Matlab:** The reviewers created a new environment using the latest versions of dependencies available at the moment of the review.
 - * **Matlab:** The code does not use external dependencies.
 - * **Python:** The reviewers used the same versions as the authors for the libraries pandas (1.5.3), numpy (1.24.3), and statsmodels (0.14.0), and created an environment with the latest compatible versions for all other dependencies.
 - * **QGIS:** The QGIS part of the package does not use external dependencies.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Figure 1A Reproduced.** The figure was manually reproduced in QGIS after opening the project file *paper_maps.qgz* and selecting the layers BRT, Standard Bus, clipped_ghs_pop, and Carto Voyager.
- **Figure 1B Reproduced.** The figure was manually reproduced in QGIS after opening the project file *paper_maps.qgz* and selecting the layers BRT, Standard Bus, Private Transit Routes, clipped_ghs_pop, and Carto Voyager.
- **Figure 1C Reproduced.** The figure was manually reproduced in QGIS after opening the project file *paper_maps.qgz* and selecting the layers Treated, Connected, Control, clipped_ghs_pop, and Carto Voyager.
- **Figure 2 Reproduced.**
- **Figure 3 Results reproduced, but the figure includes manual changes in x-axis from code output.** The exhibit was verified against the file *Fig2_tsc_pass_log.png*.
- **Figure 4 Reproduced.** The exhibits were verified against the files *fig4a_sampleAcceptRejectDotPlotunscheduled.png* and *fig4b_selection_plannedCheckInAccepted.png*.

- **Table 1** Results reproduced, but the table includes manual change for adding Average fare across all routes (₦) as last row of the table in paper. The exhibit was verified against the file *Tab1_route_comparison_BRI.tex*.
- **Table 2** Reproduced. The exhibit was verified against the file *att.tex*.
- **Table 3** Reproduced. The exhibit was verified against the file *tab3_placebo.tex*.
- **Table 4** Reproduced. The exhibit was verified against the file *tab4_driver_main_tab_MPFE_fullCont.tex*.
- **Table 5** Results reproduced, but the table includes manual change for adding one row for "p-val: Same estimate on Open", and changing the order of columns log(Fare) and log(Queues), layout of the table also differs. The exhibit was verified against the file *tab5_spillover_main_tab.tex*.
- **Table 6** Reproduced. The exhibit was verified against the file *tab6_driver_main_tab_SpillSpec_fullSpec.tex*.
- **Table 7** Reproduced. The exhibit was verified against the file *Tab7_tableWTmain.tex*.
- **Table 8** Reproduced. The exhibits was verified against the files *Tab8.tex* and *Tab8_and_S29_confidence_intervals_95.tex*.

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 903022, the Stata version used was 18.0, and the exhibits selected were: Figure S1; Table S7, Table S8, Table S10, Table S11, Table S14, Table S18, Table S24, Table S26, Table S30.

- **Figure S3** Results reproduced, the figure includes manual change in colors of the lines which does not influence the reproducibility. The exhibit was verified against the file *fig_s3_google_mob_transit_ppt.png*.
- **Table S7** Reproduced. The exhibit was verified against the file *Tab_s7_UnWeighted_means.tex*.
- **Table S8** Reproduced. The exhibit was verified against the file *tab_s8_driver_attrition_totals.tex*.
- **Table S10** Reproduced. The exhibit was verified against the file *tab_s10_driver_attrition_char.tex*.
- **Table S11** Results reproduced, the table includes manual change for adding "Number of Routes" to the paper output while not shown in code output. The exhibit was verified against the file *textitTab_S11_route_comparison_TSC_testTSTC.tex*.
- **Table S14** Reproduced. The exhibit was verified against the file *tab_s14_robustness_tab.tex*.
- **Table S18** Results reproduced, the table includes manual change for adding one row for "p-val: Same estimate on Open" , and changing the order of columns, layout of the table also differs. The exhibit was verified against the file *tab_s18_spillovers_robustness.tex*.
- **Table S24** Reproduced. The exhibit was verified against the file *TabS24_WTcheckInShift.tex*.
- **Table S26** Reproduced. The exhibit was verified against the file *tab_s26_profit_test.tex*.
- **Table S30** Reproduced. The exhibit was verified against the file *TabS30.tex*.

Reproduction Environment

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

- OS: Windows 11 Enterprise, version 23H2
- Processor: Intel(R) Xeon(R) Gold 6132 CPU @ 2.60GHz 2.60 GHz (2 processors)
- Memory available: 32 GB
- Software version:
 - Stata: StataNow 18.5 MP-Parallel Edition
 - Matlab 2024b
 - Python 3.8.19
 - R 4.4.1
 - QGIS Desktop 3.38.1