README for the Reproducibility Package for

Measuring Job Accessibility: Different Methods and New Data

Overview

This package provides information about the underlying data and codes to produce the empirical results presented in Policy Research Working Paper "Measuring Job Accessibility: Different Methods and New Data", prepared by Atsushi Iimi.

Data availability statement

The main original household data, Measuring Urban Living Standards (MULS) in Antananarivo, 2016, is publicly available. For job locations, the global datasets, such as Open Buildings 2.5D Temporal Dataset (Sirko et al., 2021), are used, which is also publicly available. Some spatial data derived from these data are computed separately by spatial software (ArcGIS). The processed data from ArcGIS are also provided in this package (spatial_data.dta). This file, along with s1_hhroster.dta (which users must obtain from obtained Measuring Urban Living Standards (MULS) in Antananarivo, 2016), allow users to execute the do-file "Tana_Access_vs_Firms_2025feb24_OpenData.do" and partially reproduce the results.

- [x] Some data cannot be made publicly available.

Data files and sources

Data files used by the Stata code:

1. s1_hhroster.dta:

Measuring Urban Living Standards (MULS) in Antananarivo, 2016

Available at https://microdata.worldbank.org/index.php/catalog/2926

2. spatial_data.dta: Derived by spatial software, using household locations and global datasets:

    Open Buildings 2.5D Temporal Dataset (Sirko et al., 2021)

    Global 30 Arc-Second Elevation (GTOPO30)

    WorldClim data, average monthly values for January during the period of 1960 to 1990

- [x] I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Input datasets used in this manuscript:

Dataset	Files	Data URL	License or license URL	Included	Notes	Accessed on
Measuring Urban Living Standards in Antananarivo 2016	s1_hhroster.dta	https://microdata.worldbank.org/index.php/catalog/2926	Direct access	No	Data is available in the data URL, but cannot be redistributed.	Nov 6, 2017
Measuring Urban Living Standards in Antananarivo 2016 – Household location	TanaLSMS2016_HH_locations_proj.shp	---	---	No	Data was collected by the team and contains personal information.	October 2017
Open Buildings 2.5D Temporal Dataset	Tana_Cell_200m_centrid.shp	http://sites.research.google/gr/open-buildings/	Creative Commons Attribution (CC BY-4.0)	Yes	Centroid created by 200m x 200m	Mar 17, 2025
Global 30 Arc-Second Elevation (GTOPO30)	elev_m.shp	https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30	https://www.usgs.gov/information-policies-and-instructions/copyrights-and-credits	Yes	Data was filtered for Antananarivo and saved in the shapefile elev_m.shp	Feb 22, 2017
WorldClim	temp_jan.shp prec_jan.shp	https://worldclim.org/	https://www.worldclim.org/about.html	No	Average monthly temperatures and precipitations for January for the city of Antananarivo were estimated for 1960-1990 and saved in the files temp_jan.shp and prec_jan.shp	Mar 17, 2025 Mar 17, 2025
Open Street Map	ALL_City_Buses.shp ALL_Suburban_Bus_OSM.shp	https://www.openstreetmap.org/	Open Data Commons Open Database License (ODbL)	Yes	Data was filtered for Antananarivo, and for taxibe bus lines Data was filtered for Antananarivo, and for suburban bus lines	Dec 22, 2018 Dec 22, 2018
City Land Use for the Master Plan Formulation for Economic Axis of TaToM (Tananarive-Toamasina, Madagascar) By the Japan International Cooperation Agency (JICA)	Job_Areas_2.shp	---	---	No	Based on city land use data in JICA TaToM city's land use map. The full report of the Master Plan Formulation is available at: https://openjicareport.jica.go.jp/pdf/12340725_01.pdf The data was shared by JICA's team working on the Master Plan Formulation. The data shared consists of the underlying data for Figure 4.1.5 in the Master Plan Formulation report	October 2018

Data processing in ArcGIS

The data processing in ArcGIS cannot be recreated because it uses data containing personally identifiable information. Below is a high-level description of the data processing. The history logs of the geospatial processing generated by ArcGIS are included in the folder ArcGIS-logs.

1. Load the households data
2. Load the temperature, precipitation, and elevation data
    a. Extract their values to points with respect of the households data
3. Load the transportation data
    b. Perform a spatial join with the household data
4. Load the jobs location data
    c. Estimate Origin-Destination matrices between the households data and the jobs location data, using the temperature, precipitation, elevation, and transportation data as inputs
    d. Extract the travel time to the closest facility for each household from the Origin-Destination matrices

Instructions for Replicators

The reproducibility package allows replicators to run the Stata do-file "Tana_Access_vs_Firms_2025feb24_OpenData.do" and reproduce figures 3-7 and tables 1-6. Replicators should follow these instructions to run the code:

1. Get access to the do-file s1_hhroster.dta.
2. Change the folder path in "Tana_Access_vs_Firms_2025feb24_OpenData.do" to the reproducibility package location
3. Run the do-file "Tana_Access_vs_Firms_2025feb24_OpenData.do".

List of exhibits

The provided code reproduces:

- [ ] All numbers provided in text in the paper

- [ ] All tables and figures in the paper

- [x] Selected tables and figures in the paper, as explained and justified below

All regression results will be reproduced, including:

- accessibility averages: Table 1
- correlation: Table 2
- summary statistics: Table 3

- probit result: Table 4
- linear IV: Table 5
- IV probit: Table 6

Figures related to descriptive statistics and regression will also be reproduced.

- distribution of commute time: figure 3
- commuting costs: figure 4
- Modal share: figure 5
- Distance to the nearest bus route: figure 6
- Travel time to CBD: figure 7

Other figures and maps are not coded but manually generated using spatial software or a geospatial web tool.

- Figure 2 is a choropleth map on population density by city area with two line layers for city bus lines traffic and for suburban bus lines traffic, for the city of Antananarivo.
- Figure 8 is a raster map (from a 200m x 200m grid obtained Open Buildings 2.5D Temporal Dataset) colored by the distance of each cell to the central business district with two line layers for city bus lines traffic and for suburban bus lines traffic, for the city of Antananarivo.
- Figure 9 is a map of the city of Antananarivo with a line layer for road connections, columns displaying the number of registered firms by city area, and commercial areas colored in red.
- Figure 10 is a screenshot of the web visualization tool for the Open Buildings 2.5D Temporal Dataset, zoomed in on Antananarivo and its central business district.
- Figure 11 is a raster map (from a 200m x 200m grid obtained Open Buildings 2.5D Temporal Dataset) colored by the area covered by buildings in each cell and a line layer for the major transportation roads.

Software Requirements

- Stata version 15