

# Replication Archive Readme File

## Road Maintenance and Local Economic Development: Evidence from Indonesia's Highways\*

Paul J. Gertler<sup>†</sup>  
*University of California, Berkeley*

Marco Gonzalez-Navarro<sup>‡</sup>  
*University of California, Berkeley*

Tadeja Gračner<sup>§</sup>  
*University of Southern California*

Alexander D. Rothenberg<sup>¶</sup>  
*Syracuse University*

July 2024

### 1 Overview

The code in this replication package for “*Road Maintenance and Local Economic Development: Evidence from Indonesia's Highways*” provides routines for building analysis datasets and producing the estimation results we present in the paper. There are two separate pieces of code contained in this replication package:

1. Code to create our analysis datasets by **processing** and **merging** different sources of data.
2. Code that uses these analysis datasets to produce the **results** presented in our paper.

**Users with Public-Use Access:** Most users of this replication archive will be individuals with only access to public-use versions of the data, which we supply. For those users, this document begins by describing how to obtain access to the confidential datasets we use (Section 2). After that, we explain how to replicate our regression results from the publicly-available analysis datasets that we supply (Section 4). Note that the `stata.zip` archive will have to be unzipped before replication proceeds.

**Users with Confidential Data Access:** Researchers who have access to the full set of confidential data files and want to replicate the entire process of building analysis datasets and producing estimation results should proceed as follows.

---

\*The authors thank seminar participants at UC Berkeley, Millennium Challenge Corporation, National University of Singapore, RAND Corporation, University of Toronto, University of Ottawa, the Joint Initiative for Latin American Experimental Economics (JILAE), and many conference participants for helpful comments and suggestions. This work was generously supported by a USAID Economic Research Partnership Grant. All errors remain our own.

<sup>†</sup>Haas School of Business, University of California-Berkeley and NBER, MC #1900 Berkeley, California 94720. Email: [gertler@haas.berkeley.edu](mailto:gertler@haas.berkeley.edu).

<sup>‡</sup>Corresponding author: Agricultural and Resource Economics, University of California-Berkeley, MC #3310, Berkeley, California 94720. Email: [marcog@berkeley.edu](mailto:marcog@berkeley.edu).

<sup>§</sup>635 Downey Way, Los Angeles, CA 90089-3332. Email: [gracner@usc.edu](mailto:gracner@usc.edu).

<sup>¶</sup>Department of Economics, Syracuse University, 426 Eggers Hall, Syracuse NY 13244-1020. Email: [adrothen@syr.edu](mailto:adrothen@syr.edu).

1. First, they users can process and organize datasets we use from raw files.
2. Then, they can merge datasets to create analysis datasets.
3. Finally, they can run the code in Section 4 to reproduce our final results.

## 2 Data Availability

**Survey and Geospatial Data from BPS:** Several datasets that we work with are confidential but are available to license and purchase directly from Indonesia’s Central Statistics Agency, *Badan Pusat Statistik*. These datasets include: (1) the Annual Census of Manufacturing Establishments (*Survei Tahunan Perusahaan Industri Pengolahan*, or SI) from 1985-2012, (2) waves of *Podes* data from 1990, 1993, 1996, 2000, 2003, 2005, and 2008; and (3) administrative boundary shapefiles. BPS’s contact information can be found below:

- Address: Jl. Dr. Sutomo 6-8 Jakarta 10710 Indonesia
- Telephone: (021) 3841195, 3842508, 3810291
- Fax: (021) 3857046
- Email: [bpshq@bps.go.id](mailto:bpshq@bps.go.id)

**Census Data:** We use unit-level census data for Indonesia for the year 2000 with an agreement with the Harvard University Library. Their contact information can be found here:

- Address: Harvard Yard; Cambridge, MA 02138, United States
- Email: [widren@fas.harvard.edu](mailto:widren@fas.harvard.edu)
- Phone: 617-495-2413

**Indonesia Family Life Survey Data:** Data from waves 1-4 of the Indonesia Family Life Survey are publically available and can be found here: <https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS.html>. In the paper, we also work with confidential information on the locations of villages. These can be found through a request to the IFLS support team. Their contact information is as follows:

- Email: [ifls-supp@rand.org](mailto:ifls-supp@rand.org)

**IPUMS International:** We work with Indonesia’s 1990 Census data as stored in IPUMS International. These data are publicly available, but to access them, you need to create an account here: <https://international.ipums.org/international/index.shtml>.

**Global Human Settlement Layer (GHSL):** The Global Human Settlement Layer data that we work with are publicly available and can be downloaded here: [https://ghsl.jrc.ec.europa.eu/ghs\\_bu.php](https://ghsl.jrc.ec.europa.eu/ghs_bu.php). We use built up area data that were constructed and calculated in [Civelli et al. \(2023\)](#).<sup>1</sup>

---

<sup>1</sup>The replication archive for this paper can be found here: <https://zenodo.org/records/7786023#.ZCwwd-zMLOQ>.

### 3 Data Processing

In this section, we describe the code for processing raw datasets and creating analysis datasets that we use for estimation.

**Computational and Software Requirements.** The code for running the processing routines were last run on an 8-core, 16GB-memory, 512GB-storage M1Pro laptop with MacOS version 11.7.10. We recommend using Stata 14 or higher for replicating this archive. The replicator should expect this code to run for about 3 hours in Stata MP 4-core.

The user also needs to run a python script to calculate shortest paths between IFLS villages and provincial capital cities. This script was last run on an 8-core, 16GB-memory, 512GB-storage M1Pro laptop with MacOS version 11.7.10 with Python 3.9.5. The routine requires installation of the `networkx` analysis package.<sup>2</sup>

**Processing Files.** The three main data processing files can be accessed here:

1. `.../IFLSroads_Replication/01_data_processing_part1.do`
2. `.../IFLSroads_Replication/02_shortestPath_desa2kota.py`
3. `.../IFLSroads_Replication/03_data_processing_part2.do`

The first and third processing Stata `.do` files call on and execute several other `.do` files. Before running `01_data_processing_part1.do` and `03_data_processing_part2.do`, you will need to change the global variable defined on line 18 of both files to reflect your local directory structure. You should only need to make this single change.

The second file is a Python script used to calculate shortest paths between IFLS villages and provincial (and district) capital cities. To run this script, you will need to change the directory paths (`filepath` and `csvpath`) on lines 13-14 to reflect your local directory structure.

Note that some of these processing steps use publicly available data, while others require access to confidential data. Some sections of this code will work correctly now, while others will require confidential data access.

The processing `.do` files are organized into several sections, as follows:

#### I. Processing IRMS Data:

- This section calls on several scripts to process, clean, and organize data from Indonesia's Integrated Road Management System (IRMS). These data contain the roughness variables used in the main analysis.
- The public raw IRMS data are stored in the following folder:  
`.../IFLSroads_Replication/Stata/IRMS/`
- Resulting data sets from these IRMS routines are contained in the following folders:  
`.../IFLSroads_Replication/Stata/dta_roads/`  
`.../IFLSroads_Replication/Stata/dta_roads/processing/`

#### II. Processing Individual IFLS Datasets:

---

<sup>2</sup>You can find more information about the `networkx` python package here: <https://networkx.org/>.

- This section calls on several scripts to clean and organize data from the first four waves of the IFLS.
- The public raw IFLS data are stored in the following folder:  
`.../IFLSroads_Replication/Stata/dta_ifls/`

### III. Processing SI Data

- This section calls on several scripts to clean and organize annual data from the SI, creating a panel of firms with harmonized geocodes, product identifiers, and analysis variables.
- The confidential raw SI data are stored in the following folder:  
`.../IFLSroads_Replication/Stata/dta_SI/`

### IV. Processing IPUMS Data

- This section runs a script that processes the 1990 IPUMS data, which we partially use in the migration analysis.
- The confidential IPUMS data are stored in the following folder:  
`.../IFLSroads_Replication/Stata/dta_IPUMS/`

### V. Organizing Road Quality and IFLS data

- This section contains scripts that calculate roughness-based travel times with road quality data, convert them into analysis datasets, and merge them into travel times information from the IFLS.
- The first 3 parts of this section are contained in  
`.../IFLSroads_Replication/01_data_processing_part1.do`
- After `01_data_processing_part1.do` is finished running, the user needs to execute a Python script to calculate roughness-based travel times between IFLS villages and provincial (and district) capital cities in Indonesia.
- The python script is available here:  
`.../Dropbox/IFLSroads_Replication/02_shortestPath_desa2kota.py`
- After this is finished running, the user can continue the data processing routine in  
`.../IFLSroads_Replication/03_data_processing_part2.do`
- The road quality data are stored in the following folder:  
`.../IFLSroads_Replication/Stata/dta_roads/`
- The transportation network and travel time data are stored in the following folder:  
`.../IFLSroads_Replication/Stata/dta_transportNetworks/`

### VI. Processing GDRP Data

- This is a series of scripts that processes regional gross domestic product data. These data were scanned and digitized into `.csv` files from a series of hardcopy publications that were produced by BPS.
- The raw `.csv` files, which are publicly available, can be found here:  
`.../IFLSroads_Replication/Stata/csv/GDRP/`

### VII. Additional Datasets

- This section carries out scripts to process several additional datasets, including:
  - (A) province adjacency and GDP for adjacent provinces

- (B) calculating budget IVs
- (C) processing publicly available election data
- (D) oil, gas, and mining shares (from DAPOER);
- (E) Built Up Area data from GHSL
- (F) FDI shares from the SI
- (G) DPU budget data from hardcopies
- (H) PODES data
- (I) hausman, and island market potential variables.

## VIII. Merge Processing

- This section runs scripts that carry out merges for our main analysis datasets.

## IX. Prepping Analysis Datasets

- This section runs scripts that create our main analysis datasets which exist for different units of analysis.

## X. Scrambling Identifiers

- The final part of this script scrambles district, household, and community identifiers to produce the publicly available replication datasets we supply.

## 4 Running the Main Analysis

In this section, we describe the code for conducting the empirical analysis that we present in our paper. This analysis can be conducted from the public-use versions of the data that we supply. Note that these public-use versions contain aggregates of administrative road quality data, make use of confidential IFLS geocodes for merging, and contain aggregates of several other variables that use non-public, confidential data from BPS. As a result, we have scrambled the district and province-level identifiers in the public use versions to preserve confidentiality.

The `04_main_analysis.do` file runs all of the code on the public-use analysis datasets to generate the estimates and graphs/maps for the 25 tables (18 in the Appendix) and 18 figures (13 in the Appendix) in the paper.

**Computational and Software Requirements.** The code for running the analysis routines were last run on an 8-core, 16GB-memory, 512GB-storage M1Pro laptop with MacOS version 11.7.10. We recommend using Stata 14 or higher for replicating this archive. The replicator should expect this code to run for about 3 hours for analysis in Stata MP 4-core.

**Folder Information.** This replication Stata folder `.../IFLSroads_Replication/Stata/` includes the following sub-directories:

- `/csv/`: contains several csv files used in replication
- `/do/Analysis/`: contains all of the `.do` files used in the analysis which produce the tables and figures for the paper.
- `/do/consumption_dofiles/`: contains `.do` files used to process consumption aggregates and prices from the IFLS data
- `/do/GDRP/`: contains `.do` files used to process the GDRP `.csv` files and assemble them into a merged datasets for use in replication
- `/do/IFLS`: contains `.do` files used to process the raw IFLS data
- `/do/processing`: stores scripts for processing datasets into main analysis datasets
- `/do/roadsData`: contains a `.do` file to process road quality data and extract network information
- `/do/SI`: contains several `.do` files for processing the SI data
- `/do/travelTimes`: contains `.do` files that process and organize travel times data from the IFLS and merge them to travel times data from the IRMS
- `/dta/`: containing all `dta` files needed for analysis
- `/dta_ifls`: containing all `.dta` files from the IFLS
- `/dta_IPUMS`: contains `.dta` files derived from IPUMS data
- `/dta_PODES`: contains `.dta` files derived from PODES data
- `/dta_roads`: contains `.dta` files derived from the IRMS data
- `/dta_SI`: contains `.dta` files derived from the SI data

- /dta\_transportNetworks: contains .dta files derived from the transport network data

### Replication Instructions.

- The main do file for running the analysis can be found here:

```
.../04_main_analysis.do
```

This file calls on several subsequent .do files, also stored in the /do/Analysis/ folder, that each carry out separate pieces of the analysis we present in the paper. Before running the 04\_main\_analysis.do file, you will need to change the global variable defined on line 16 to reflect your local directory structure. You should only need to make this single change.

- For our main tables, figures, appendix tables and appendix figures, a detailed list of the individual, table-specific .do files and the analysis results they create are described below.
- The results are all stored in the .../Writeup/tabFig/ folder.
- After the Stata routines are finished, a file in the repository called .../Writeup/Replication\_tabFig.tex can be compiled to reproduce the tables and figures in our paper.

### Main Tables.

- **Table 1:** run using /do/Analysis/04\_c\_budgetIV\_regressions.do
- **Table 2:** run using /do/Analysis/05\_a\_Table2\_IFLS\_indiv\_V3\_newIVs.do
- **Table 3:** run using /do/Analysis/06\_a\_Table3\_kabuYear\_newIVs.do
- **Table 4:** run using /do/Analysis/07\_b\_Table4\_SI\_indiv\_newIVs.do
- **Table 5:** run using /do/Analysis/08\_a\_Table5\_comm\_prices\_newIVs.do
- **Table 6:** run using /do/Analysis/09\_a\_Table6\_migration\_newIVs.do
- **Table 7:** run using /do/Analysis/11\_c\_welfareDecomp.do

### Main Figures.

- **Figure 1:** run using /do/Analysis/15\_roadQuality\_graphs.do
- **Figure 2:** run using /do/Analysis/16\_tcost\_maps.do
- **Figure 3:** run using /do/Analysis/15\_roadQuality\_graphs.do
- **Figure 4:** run using /do/Analysis/11\_d\_costBenefit.do
- **Figure 5:** run using /do/Analysis/11\_d\_costBenefit.do

### Appendix Tables.

- **Table A.1:** run using `/do/Analysis/10_a_Table1_travelTimes_newIVs.do`
- **Table A.2:** run using `/do/Analysis/04_b_budget_compare_DPU.do`
- **Table A.3:** run using `/do/Analysis/04_a_firstStage_table.do`
- **Table A.4:** run using `/do/Analysis/04_c_budgetIV_regressions.do`
- **Table A.5:** run using `/do/Analysis/05_c_Table2_IFLS_indiv_V3_newIVs_nonMovers.do`
- **Table A.6:** run using `/do/Analysis/05_e_Table2_IFLS_indiv_V3_newIVs_example.do`
- **Table A.7:** run using `/do/Analysis/06_d_Table3_kabuYear_newIVs_numFirms.do`
- **Table A.8:** run using `/do/Analysis/17_hedonics.do`
- **Table A.9:** run using `/do/Analysis/08_b_Table5_comm_prices_newIVs_hedonics.do`
- **Table A.10:** run using several `.do` files:
  - Panel A: `/do/Analysis/10_d_Table1_travelTimes_newIVs_IVrobust.do`
  - Panels B and C: `/do/Analysis/05_d_Table2_IFLS_indiv_V3_newIVs_IVrobust.do`
  - Panel D: `/do/Analysis/06_c_Table3_kabuYear_newIVs_IVrobust.do`
- **Table A.11:** run using several `.do` files:
  - Panel A: `/do/Analysis/10_b_Table1_travelTimes_newIVs_robust.do`
  - Panels B and C: `/do/Analysis/05_b_Table2_IFLS_indiv_V3_newIVs_robust.do`
  - Panel D: `/do/Analysis/05_b_Table2_IFLS_indiv_V3_newIVs_robust.do`
- **Table A.12:** run using several `.do` files:
  - Panel A: `/do/Analysis/10_d_Table1_travelTimes_newIVs_IVrobust.do`
  - Panels B and C: `/do/Analysis/05_d_Table2_IFLS_indiv_V3_newIVs_IVrobust.do`
  - Panel D: `/do/Analysis/06_c_Table3_kabuYear_newIVs_IVrobust.do`
- **Table A.13:** run using several `.do` files:
  - Panel A: `/do/Analysis/10_d_Table1_travelTimes_newIVs_IVrobust.do`
  - Panels B and C: `/do/Analysis/05_d_Table2_IFLS_indiv_V3_newIVs_IVrobust.do`
  - Panel D: `/do/Analysis/06_c_Table3_kabuYear_newIVs_IVrobust.do`
- **Table A.14:** run using several `.do` files:
  - Panel A: `/do/Analysis/10_e_Table1_travelTimes_newIVs_clstrRobust.do`
  - Panels B and C: `/do/Analysis/05_f_Table2_IFLS_indiv_V3_newIVs_clstrRobust.do`
  - Panel D: `/do/Analysis/06_f_Table3_kabuYear_newIVs_clstrRobust.do`
- **Table A.15:** run using `/do/Analysis/06_e_Table3_kabuYear_newIVs_adjacent.do`
- **Table A.16:** run using `/do/Analysis/11_b_welfareDecomp_expShares.do`
- **Table A.17:** run using `/do/Analysis/11_b_welfareDecomp_expShares.do`

- **Table A.18:** run using `/do/Analysis/11_a_deterioration.do`

## Appendix Figures.

- **Figure A.1:** run using `/do/Analysis/15_roadQuality_graphs.do`
- **Figure A.2:** run using `/do/Analysis/15_roadQuality_graphs.do`
- **Figure A.3:** run using `/do/Analysis/14_desaMaps.do`
  - Note that this figure cannot be replicated without access to confidential IFLS location identifiers.
- **Figure A.4:** run using `/do/Analysis/04_b_budget_compare_DPU.do`
- **Figure A.5:** *Source:* [World Bank \(2012\)](#)
- **Figure A.6:** *Source:* [World Bank \(2012\)](#)
- **Figure A.7:** *Source:* [World Bank \(2008\)](#)
- **Figure A.8:** *Source:* [World Bank \(2008\)](#)
- **Figure A.9:** *Source:* [Ahmad and Mansoor \(2002\)](#)
- **Figure A.10:** run using `/do/Analysis/13_roadLength.do`
- **Figure A.11:** run using `/do/Analysis/11_a_deterioration.do`
- **Figure A.12:** run using `/do/Analysis/11_d_costBenefit.do`

## References

- AHMAD, E. AND A. M. MANSOOR (2002): *Indonesia: Managing Decentralization*, International Monetary Fund.
- CIVELLI, A., A. GADUH, A. D. ROTHENBERG, AND Y. WANG (2023): "Urban Sprawl and Social Capital: Evidence from Indonesian Cities," *The Economic Journal*, 133, 2110–2146.
- WORLD BANK (2008): *Spending for Development: Making the Most of Indonesia's New Opportunities*, World Bank.
- (2012): *Investing in Indonesia's Roads: Improving Efficiency and Closing the Financing Gap*, World Bank.