Data and Code for: The Origin of the State: Land Productivity or Appropriability? Comment

Nikolai Cook, Wilfrid Laurier University

Thibaut Duprey, Bank of Canada

Anthony Heyes, University of Birmingham

Martino Pelli, Asian Development Bank

OVERVIEW

The code in this replication package constructs the figure and tables from a single data source (Mayshar et al. 2022) using Stata.

A single do file "do file v2.do" runs all of the code for all tables and figures.

The dataset "DATASET_ETHNOATLAS.dta" from Mayshar et al 2022 is included in this replication package.

The Replicator should expect the code to run for about 10 seconds on a "standard" desktop.

DATA AVAILABILITY AND PROVENANCE

The data is from Mayshar et al. 2022 in unaltered form, is provided in the replication package and can also be downloaded at the journal website:
https://www.journals.uchicago.edu/doi/suppl/10.1086/718372

STATEMENT ABOUT RIGHTS

I certify that the authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.

SUMMARY OF AVAILABILITY

All data are publicly available.

DETAILS ON EACH DATA SOURCE

Data.Name	Data.Files	Location	Provided	Citation
DATASET_ETHNOATLAS.dta	DATASET_ETHNOATLAS.dta	Root Folder	True	Mayshar et al. 2022

A copy of the data is provided as part of this archive. Data in this manuscript was downloaded from

https://www.journals.uchicago.edu/doi/suppl/10.1086/718372

COMPUTATIONAL REQUIREMENTS

We used Stata 19 for the analysis, but earlier versions should be compatible. The do file we provide also installs the required packages "estout, ivreg2, ranktest, pdslasso, lassopack, and winsor"

SOFTWARE REQUIREMENTS

We used Stata 19 for the analysis, but earlier versions should be compatible. The do file we provide also installs the required packages "estout, ivreg2, ranktest, pdslasso, lassopack, and winsor"

CONTROLLED RANDOMNESS

None

MEMORY, RUNTIME, STORAGE REQUIREMENTS

Dataset is less than 1MB in total.

SUMMARY

Approximate time needed to reproduce the analysis on a standard 2025 desktop machine is 1 minutes.

DETAILS

The code was last run on a mac with an Apple M2 Ultra and 128GB of RAM running MacOS 14.5 was used.

DESCRIPTIONS OF PROGRAMS/CODE

A single do file "do file v2.do" is used.

Open it and press "run."

INSTRUCTIONS TO REPLICATORS

1. Ensure that you have "DATASET_ETHNOATLAS.dta" and the "do file v2.do" files in the appropriate directories. This can be done by ensuring the dataset and the do file are in the same directory.

2. Double click that do file, which will make its location the working directory. Now hit "run".

3. Wait approximately 30 seconds for all packages to be downloaded, and for all tables and figures to be produced, in the same directory.

DETAILS

"do file v2.do" will do everything.
Please note that a datasignature assert command is in the do file, to ensure the same dataset (which was used in Mayshar et al (2022) and this replication) is being used.

A partial codebook to allow for ensuring the ethnographic atlas data is comparable:
V33 is the main dependent variable, Jurisdictional Hierarchy.
V29 is the primary food source, with 6 being cereals.
V5 is dependence on agriculture.

LIST OF TABLES AND PROGRAMS

The provided code reproduces all tables and figures in the paper.

Figure/Table	Program	Line	Output File	Note
Fig 1	"do file.do"	200	fig_1.png	
Tab 1	"do file.do"	223	table_1.tex	
Tab 2	"do file.do"	259	table_2.tex	
Tab A1	"do file.do"	325	table_a1.tex	

References

Mayshar, J., Moav, O., & Pascali, L. (2022). The origin of the state: Land productivity or appropriability?. *Journal of Political Economy*, *130*(4), 1091-1144.