README File on Data and Codes for

"Land Security and Mobility Frictions"

By Tasso Adamopoulos, Loren Brandt, Chaoran Chen, Diego Restuccia, and Xiaoyun Wei

Quarterly Journal of Economics

Data Availability Statement:

Our main source of micro data is the National Fixed Point Survey carried out by the Research Center for Rural Economy (RCRE, 2004), Chinese Ministry of Agriculture. Earlier years of this data set have been used in the literature, see for instance Chari et al. (2021) and Adamopoulos et al. (2022), who provide excellent descriptions of this dataset from different perspectives. The survey was first conducted in 1986, but it was only in 2003 that individual-level labor supply data were collected. We focus on the period 2004-2018 to exploit the novel individual-level data. We accessed the data through Shanghai Jiao Tong University, with which one of our collaborators was affiliated when we worked on this paper. Shanghai Jiao Tong University has purchased collective access to this data source from the RCRE. We accessed the data most recently on October 7, 2023.

We note that the raw data are not available for public download. We provide all the Stata files used in cleaning and constructing the data moments used in our analysis. We also provide data files with the data moments that are used in our analysis. Conditional on these data moments, the remaining files can be used to replicate all the results.

We also use the statistical yearbook (National Bureau of Statistics of China, 2005) to calculate sectoral value-added shares and employment shares. This data source is publicly available from NBS's website. We use the 2005-2019 yearbooks that include information for our sample period 2004-2018.

Computational Requirements:

- Codes in the folder "Stata-Files" require Stata 18.0 (and the access to the raw data files).
- Codes in the folder "Matlab-Files" require Matlab 2022b.

- All Matlab programs are run on a workstation that features AMD Ryzen 3970x CPU with 32 cores 64 threads and 128GB memory.

File Structure:

- "Replication-Files/Stata-Files": in total 5 folders
  - "dofiles" (2 do-files)
  - "results" (location for Stata results)
  - "temp-data" (storing temporary files to be deleted when the code ends)
  - "working-data" (storing data files after cleaning)
- "Replication-Files/Data-Statistics": in total 3 Excel files.
  - "Calibration.xlsx" (data statistics calculated by the Stata codes that are used in Tables I, II, and VIII in the main text, and Tables A.1 and A.2 in the online appendix, as well as used in the model calibration)
  - "Calibration-Remote.xlsx" and "Calibration-Suburban.xlsx" (same as "Calibration.xlsx" but for remote and peri-urban areas, respectively)
  - "Dummy-Nonagr.xlsx" and "Operator.xlsx" (results for Appendix E)
  - "Accounting.xlsx" (results for Appendix A.3)
- "Replication-Files/Matlab-Files": in total 2 folders and 2 m-files
  - "matfiles" (storing Matlab data files including data moments and model statistics that are used by Matlab subroutines)
  - "subroutines" (some m-files to be called by the main m-file)
  - "main.m" (main m-file that calculates model statistics for the baseline analysis in the paper, including Tables V, VI, VII, IX, X, and XI in the main text, and Tables B.5, C.7, C.8, F.9, and G.10 in the online appendix)
  - "calibrationOPTIONAL.m" (m-file that performs calibration for various settings, with the comparison between model and data moments and listings of model parameters)

Steps to Run the Files:

Step 1. [OPTIONAL] The Stata codes in the "Stata-Files" folder cannot be run without raw data files. If you have access to the raw data files, then running these files will generate data moments as the Excel files in the folder "Data-Statistics".

1a. Run the code "data-cleaning.do". This code cleans the raw data and then generate two data files called "basic-cleaning-done.dta" and "ready-for-analysis-individual.dta" in the folder "temp-data" that will be used in the next step.

1b. Run the code "compute-moments.do". This code computes data moments used in our calibration, stored in the Excel file "Calibration.xlsx". This code also generates results for Appendix E that are stored in "Operator.xlsx" and "Dummy-Nonagr.xlsx". Note that if you wish to obtain the results for remote versus peri-urban areas, uncomment the lines 675-686.

Step 2. Run the code "main.m" in the folder "Matlab-files". This is the main code that computes results for the quantitative analysis. This code shows parameter values for our model and the comparison between model and data moments as results of the calibration process. Note that the parameter values and the data moments have been stored to the mat-files which are read by this code. This code also performs counterfactual experiments by eliminating land insecurity and labor mobility barriers separately, for various settings (baseline, extensions, and robustness). Results will be displayed on your screen with clear labels referring to the sections/tables of the paper. Specifically, this code generates results for Tables V, VI, VII, IX, X, and XI in the main text, and Tables B.5, C.7, C.8, F.9, and G.10 in the online appendix, as well as some numbers mentioned in the text in Section V.

Step 3. [OPTIONAL] If you are interested in the calibration procedure, you can run the code "calibrationOPTIONAL.m". This code searches for parameter values that minimize the distance between model moments and data moments, for various settings (baseline, extensions, and robustness) and then stores the parameter values in mat-files. Note that this code takes a very long time (see the next section) to run on a typical personal computer, and hence we have already stored the calibrated parameter values in mat-files that have been used in Step 2. The particle swarm algorithm needs initial values. We have made the initial values equal to the chosen parameter values and hence this code is somewhat for illustrative purposes. After finding the minimum, the code also displays the resulted parameter values as well as the comparison between model and data moments.

Computational Time:

- The Stata codes take around 2 hours to finish, with most of the time spent on linking individuals over time to construct a panel data and on bootstrap repetitions to obtain the standard errors in the Excel files.
- The code "main.m" takes 3 minutes to finish on a workstation of AMD Ryzen 3970x with 32 cores 64 threads, 128GB memory. We also tried on a MacBook Pro 16 inches with Intel Core i7-9750H (6 cores 12 threads) and it takes about 6 minutes.
- The code "calibrationOPTIONAL.m" takes substantially longer. Specifically, this code benefits substantially from parallel computation. We allow for parallel computation on our workstation of AMD Ryzen 3970x with 32 cores 64 threads, 128GB memory, and it takes around 12 hours. We have not tried it on the aforementioned MacBook Pro; we expect it would take roughly a week to finish. Again, you do not need to run this code on your personal computer. In Step 2, we directly load into the program the stored parameter values from the calibration process as well as the comparison between model and data moments.

Contact:

Please contact Chaoran Chen at chenecon@yorku.ca if you need any further information. We will try to help you in reasonable replication attempts within five years after the publication of this article.

References:

- Adamopoulos, T., Brandt, L., Leight, J., and Restuccia, D. (2022), "Misallocation, Selection, and Productivity: A Quantitative Analysis with Panel Data from China," *Econometrica*, 90: 1261-1282.
- Chari, A., Liu, E. M., Wang, S.-Y., and Wang, Y. (2021), "Property Rights, Land Misallocation, and Agricultural Efficiency in China," *The Review of Economic Studies*, 88(4): 1831-1862.
- National Bureau of Statistics of China (2005), "China Statistical Yearbook 2005-2019, [dataset]," publicly available at https://www.stats.gov.cn/sj/ndsj/.
- Research Center for Rural Economy (2004), "National Fixed Point Survey [dataset]," unpublished data.