Minki Kim and Munseob Lee, ”The U.S. Structural Transformation and Regional Convergence: Racial Heterogeneity,”
Journal of Applied Econometrics

1 Overview

The code in this package replicates all figures and tables in Kim and Lee (Forthcoming).

2 Data Availability and Source

The paper uses the U.S. census microdata from 1940 to 2020 provided by IPUMS-USA (Ruggles et al., 2023). IPUMS-USA does not allow for redistribution without permission. To replicate the results, one should download the raw data from the IPUMS-USA website. The exact samples used for each year are as follows:

- 1940 1% sample
- 1950 1% sample
- 1960 5% sample
- 1970 1% state fm1
- 1980 1% metro
- 1990 1% unweighted state sample
- 2000 5% sample
- 2010 ACS
- 2020 ACS 5yr

Variables needed to run the code are as follows:

- YEAR: Census year
- PERWT: Person weight
- REGION: Census region and division
- STATEICP: State (ICPSR code)
- AGE: Age

- RACE (general): Race [general version]
- BPL (general): Birthplace [general version]
- EMPSTAT (general): Employment status [general version]
- IND1950: Industry, 1950 basis
- WKSWORK2: Weeks worked last year, intervalled
- INCWAGE: Wage and salary income
- MIGRATE5 (general): Migration status, 5 years [general version]
- MIGRATE1 (general): Migration status, 1 year [general version]

All variables are harmonized variables across all sample years.

3 Statement about Rights

We certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

4 Software and Memory Requirement

The code was last run on a workstation with 13th Gen Intel(R) Core(TM) i9-13900, 64GB RAM, Windows 11 Enterprise with Stata/MP 18.0, each file took the following amount of time:

- Figure1.do: 70 seconds
- Table1.do: 34 seconds
- Table2.do: 75 seconds
- Table3.do: 72 seconds
- TableA1.do: 85 seconds
- TableA2.do: 79 seconds
- TableA3.do: 72 seconds

The code is compatible with Stata version 16 or 17. Earlier versions of Stata are not tested.

5 Decription of programs / code

All .do files are stored in the "code" folder. All outputs (figures in pdf, excel spreadsheets and .dta files for tables and figures) are stored in the "output" folder.

- `DataClearing.do` and `DataClearing_A2.do` take the raw IPUMS dataset, impose sample restrictions (see Section 2 and Online Appendix for details), define four regions, calculate aggregated statistics, and store them into a single dataset, `temp_composite.dta` in the `data/proc` directory. The dataset `temp_composite.dta` contains the following aggregate variables:
  - `year`
  - `region`: dummy for US South, North, Midwest, and West
  - `sector`: dummy for agriculture and non-agriculture
  - `wage`: Year-region-sector level average wage, corresponding to w^i_{jt}
  - `agri`: Employment share of agriculture in region i and year t
  - `nonagri`: Employment share of non-agriculture in region i and year t
  - `agri_yr`: Employment share of agriculture in year t
  - `nonagri_yr`: Employment share of nonagriculture in year t
  - `north1/midwest1/south1/west1`: North/Midwest/South/West share of agricultural employment in year t
  - `north2/midwest2/south2/west2`: North/Midwest/South/West share of nonagricultural employment in year t
  - `north/midwest/south/west`: Employment share of North/Midwest /South/West in year t
- `Decomposition.do` takes the `temp_composite.dta` file and run the decompostion analysis. Specifically, the code constructs each variable in Equation 3 in the paper and calculates the numbers in Table 2, 3, A1, A2, and A3, depending on the specifications.
- `Figure1.do` uses `DataClearing.do` and `Decomposition.do` and generate Figure 1 in the paper.
- `Table1.do/Table2.do/Table3.do/TableA1.do/TableA3.do` use `DataClearing.do` and `Decomposition.do` and generate Table 2, 3, A1, and A3 in the paper. `TableA2.do` uses `DataClearing_A2.do` and `Decomposition.do` generates Table A2 in the paper.

6 Instructions to Replicators

- Unzip the replication package. Ensure the folder structure is in order. There should be three folders: `code`, `data`, and `output`. The `data` directory has two sub-folders: `proc` and `raw`, which are initially empty. The `code` folder contains 10 `do` files and the `output` folder contains 14 items (2 items for each figure and table).
- All codes are written and run in Stata. No additional package installation is required.
- Download the raw dataset from IPUMS-USA as referenced above in a Stata `dta` format and put it under the `data/raw` directory as `census1940_2020_raw.dta`. The raw dataset contains 51,601,093 observations. The number of observations for each sample are as follows: The exact samples used for each year are as follows:

  - 1940 1% sample: 1,351,732
  - 1950 1% sample: 1,922,198
  - 1960 5% sample: 8,965,606
  - 1970 1% state fm1: 2,030,386
  - 1980 1% metro: 2,267,320
  - 1990 1% unweighted state sample: 2,479,020
  - 2000 5% sample: 14,081,466
  - 2010 ACS: 3,061,692
  - 2020 ACS 5yr: 15,441,673

- The default path is `C:/replication package`. If your path is different, adjust the path at the top part of `Figure1.do`, `Table1.do`, `Table2.do`, `Table3.do`, `TableA1.do`, `TableA2.do`, and `TableA3.do`. There is no need to edit `DataClearing.do`, `DataClearing_A2.do` and `Decomposition.do` files, because those codes are only called from other codes and not directly run.
- When run, all `.do` files automatically delete the interim `.dta` files and only save the final output. Comment out the last section of each `.do` files if you would like to keep the interim data files. Unlike the final outputs, they are stored in `data/proc` directory.
- Run `Figure1.do`, `Table1.do`, `Table2.do`, `Table3.do`, `TableA1.do`, `TableA2.do`, and `TableA3.do` to generate all figures and tables in the paper. There is no sequence. Each code runs indepedently.

References

KIM, M. AND M. LEE (Forthcoming): "The US Structural Transformation and Regional Convergence: Racial Heterogeneity," *Journal of Applied Econometrics*.

RUGGLES, S., S. FLOOD, M. SOBEK, C. DANIKA, BROCKMAN GRACE, S. RICHARDS, AND M. SCHOUWEILER (2023): "IPUMS USA: Version 13.0," *Minneapolis, MN: IPUMS*, https://doi.org/10.18128/D010.V13.0.