

1 Computational Requirements

- Software Requirements
- Computational Requirements
- Time Requirements
- Package Installation

2 Instructions

2.1 Overview

As of July 12th, 2023 the file `master.do` contains the code to create the two final datasets, `cz_pooled.dta` and `cz_stacked.dta`. To be precise, it sets up and runs the subfiles in the correct order to create those datasets. To run it, you need to add the absolute paths on your computer to the if-else block at the top of the file as I have for my (Everett Stamm's) computer. You will need a path to the dropbox folder where the data is stored, a path to the repository where the code is stored, a path to your FFMPEG installation (only necessary for map creation, not data cleaning), a path to your Rterm.exe installation, and a flag for if you will use the `gz7` function.

2.2 gz7 setup

As of 05.17.2022, the `gzuse` package no longer works on Windows computers. To get around this, I created the wrapper function `gz7` that calls `7zip` to unzip the file. To set this up, please install `7zip` and add it to your system path. <https://www.7-zip.org/download.html>

2.3 R setup

You will need the following R packages. Please install them before running the code:

- `tidyverse`
- `sf`
- `haven`
- `tigris`
- `stringr`
- `readxl`

2.4 Derenoncourt setup

The entirety of the Derenoncourt "Can You Move to Opportunity? Evidence from the Great Migration." replication package is in the dropbox in `"/municipality_proliferation/derenoncourt_opportunity/replication_AER/`. Within this, the folder `"code"` is her original replication code and the folder `"code_replication"` replicates it to our specifications. To make things easier, I've copied the files with major changes into the repository. All of the `do-files` starting with `A` are new and used to create new data for the decades stacked sample. `4_final_dataset.do` and `4_final_dataset_split.do` are modified versions of Derenoncourt's similarly named file that only includes variables we need and modifies it in some other ways for our analysis (e.g. raw instruments instead of rank). To slim things down, I've copied the raw derenoncourt data that we need into the dropbox folder `"/municipality_proliferation/raw/dcourt/"` and the modified code we need in the Github as `"code/dcourt_setup/"`.

The changes to the pooled sample are minimal, only dropping unused data and adding some new data. The stacked sample requires adding 1950 and 1960 data, but largely follows the same logic. The system of copying data from the Derenoncourt replication package and cleaning/analyzing it with a modified version of her code feels a bit ad-hoc to me, I would consider rethinking that bit of architecture. Or just say that if it ain't broke, don't fix it. Your choice.

The code also creates `cz_pooled_south.dta` and `cz_stacked_south.dta`. These are similar versions of `cz_pooled.dta` and `cz_stacked.dta`, with the addition of several new versions of the instrument: using rural migrants only (`rm`), considering texas as a northern state (`nt`), and allowing southern cities to be destinations (`sc`). The original Derenoncourt data was created on a proprietary sample from the 1940 full count census (`municipality_proliferation/derenoncourt_opportunity/replication_AER/population/raw/170515_popc1940.dta`) to create 1940 city population values that, for some reason or another, I have not been able to exactly replicate from the 1940 full count census that's on IPUMS. We needed to include Texan and southern cities for these versions of the instrument, so I had to use what was on IPUMS, thus you'll see the original version of the instrument that's in the `_south` datasets is slightly different than the original datasets.

2.5 Code overview

- code/ado/

appendmodels.ado Ben Jann's appendmodels function to create nice tables stacking models in rows, from <https://www.stata.com/11/msg01289.html>

cityfix_ccdb.ado Wrapper for all the city name string cleaning to harmonize city names in CCDB data.

cityfix_census.ado Wrapper for all the city name string cleaning to harmonize city names in Census data.

gz7.ado Proprietary function to unzip GZIP files, since evidently the SSC package gzuse no longer works on Windows. Wouldn't be surprised if they update gzuse to fix it eventually and make this irrelevant.

- code/analysis/

alt_inst_tables.do Creates tables of first stage, OLS, reduced form, and 2SLS using the baseline, Residualized state FE, top urban dropped, 1940 southern state of birth, european migrant, and southern white migrant instruments, as well as the baseline instrument with outcomes expressed in per capita of total population (as opposed to urban population).

alt_inst_tests.do Creates our version of figure D16 from the Derenoncourt online appendix, the Hansen J-tes. Also creates figure D14, binscatters of the reduced form for the european and southern white instruments.

balancetables.do Creates balance tables, regressing 1940 baseline characteristics on the instrument, and then creates tables comparing the first stage, OLS, reduced form, and 2SLS with and without the baseline controls deemed significant. Also creates combined balance tables, regressing the instrument on all the baseline variables at once.

binscatter.do Creates various binscatter graphs of the instrument.

loo_text.do Performs the leave-one-out tests.

origin_shocks_dest_chars.do Creates dataset of origin county shocks and the characteristics of their primary destination city. Not currently used for anything, just informative.

placebo_test.do Performs placebo tests.

rmscnt_inst_tables.do Creates tables of first stage, OLS, reduced form, and 2SLS using the baseline instrument and the instruments created using rural migrants only (rm), considering texas as a northern state (nt), allowing southern cities to be destinations (sc), and all permutations of them.

- code/cleaning/

census_race_cleaning.do Creates measures of total population by race for each decade from 1910-2020.

census_urban_populations.do Creates measures of urban population by race for each decade from 1900-1930 and measures of the fraction of population in the main city in a CZ.

cgoodman_place_county_geog.R Creates measures of fraction land incorporated from the Chris Goodman data.

cog_cleaning.do Cleans the census of governments data.

covariates.R Creates 1940 Baseline covariates.

cz_court_orders.do Links Chris Reardon school desegregation court orders to CZs.

dataprep.do Merges all data and creates cz_pooled.dta, cz_stacked.dta, cz_pooled_south.dta, and cz_stacked_south.dta

dcourt_cleaning Creates total population versions of Derenoncourt instruments/data. NOTE: this needs to be either updated or removed, we don't seem as interested in total population instruments and this needs to be updated to reflect that we moved the Derenoncourt data from it's own folder to be part of our code structure.

geogs.do Finalizes fraction incorporated and fraction unusable from cgoodman_place_county_geog.R and lutz_sand_cleaning.do.

lutz_sand_cleaning.do Cleans the Lutz and Sand measures of buildable land.

- code/dcourt_setup/

2_lasso.do Performs lasso variable selection.

3_instrument.do Creates all our shift-share instruments.

4_final_dataset.do Creates final pooled dataset from Derenoncourt (modified) with all instruments and variables sourced from her data.

4_final_dataset.do Creates final stacked dataset from Derenoncourt (modified) with all instruments and variables sourced from her data.

A1_census_1950_1960_racepop.do Creates urban race population data for 1950 and 1960 from IPUMS samples.

A2_clean_cz_mobility_1900_2015.do Creates mobility dataset.

A4_clean_city_population_census_1940_full.do Creates urban race population data for 1940 from IPUMS full count census.

A5_clean_cz_snq_european_immigration_instrument.do Creates urban race population data for 1940 from IPUMS full count census.

- code/helper/

bartik_generic.do Code to create Bartik Shift-Share instrument. Adopted from Derenoncourt code.

3 Data Sources (INCOMPLETE)

3.1 Derenoncourt 2022

Availability:

Usage: Users should download the full repository into `/municipality_proliferation/derenoncourt_opportunity/` and then follow the instructions in ReadMe.pdf to acquire the necessary data. Note that none of the data listed as unavailable is required for our analysis.

Citation: Ellora Derenoncourt. Can you move to opportunity? evidence from the great migration. *American Economic Review*, 112(2):369–408, 2022. doi: <https://doi.org/10.1257/aer.20200002>

3.2 IPUMS USA

Availability: Extracts permitted for replication purposes.

Citation: Steven Ruggles, Sarah Flood, Matthew Sobek, Danika Brockman, Grace Cooper, Stephanie Richards, and Megan Schouweiler. IPUMS USA: Version 13.0. [dataset], 2023. URL <https://doi.org/10.18128/D010.V13.0>

3.3 NHGIS

Availability: Extracts permitted for replication purposes.

Citation: Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 17.0. [dataset], 2022. URL <http://doi.org/10.18128/D050.V17.0>

3.4 US Census Codes

Availability: Public

Citation: US Census Bureau. American National Standards Institute, Federal Information Processing Series, and Other Standardized Geographic Codes. [dataset], 2023. URL <https://www.census.gov/library/reference/code-lists/ansi.html>

References

US Census Bureau. American National Standards Institute, Federal Information Processing Series, and Other Standardized Geographic Codes. [dataset], 2023. URL <https://www.census.gov/library/reference/code-lists/ansi.html>.

Ellora Derenoncourt. Can you move to opportunity? evidence from the great migration. *American Economic Review*, 112(2):369–408, 2022. doi: <https://doi.org/10.1257/aer.20200002>.

Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 17.0. [dataset], 2022. URL <http://doi.org/10.18128/D050.V17.0>.

Steven Ruggles, Sarah Flood, Matthew Sobek, Danika Brockman, Grace Cooper, Stephanie Richards, and Megan Schouweiler. IPUMS USA: Version 13.0. [dataset], 2023. URL <https://doi.org/10.18128/D010.V13.0>.