## Overview

The material in this replication package is for the paper **Spatial Unit Roots and Spurious Regression** by Ulrich Müller and Mark Watson.

There are two types of data:

* Data from Chetty, Hendren, Kline and Saez (2014).
* Simulated data generated by programs included in this package.

The data from Chetty, Hendren, Kline and Saez are processed using Matlab and Stata. The simulated data are generated using Fortran and Mathematica. Approximate runtimes are provided below.

## Data

The data are taken from the replication materials for Chetty, Hendren, Kline and Saez (2014).

The replication materials for the Chetty et al paper can be found here: <https://opportunityinsights.org/data/>

The data used in our paper is in the Chetty et al file called **online\_data\_tables.xls**.

We reformatted the data from **online\_data\_tables.xls**, added location information and saved the resulting data in the file **Chetty\_Data\_1.xlsx**

In this replication package, the data processing files are in the folder **Data\Chetty\_Mobility\**. There you can find a copy of **Chetty\_Data\_1.xlsx** along with other files used to process the data.

Details:

In the **Chetty\_Data\_1.xlsx** file

* The variables in columns M-AT (*FracBlack* through *IncGrth*) were copied from tab “Online Data Table 8” in **online\_data\_tables.xls**
* The variables in columns G-L (*AM* through *RMMale*) were taken from tab “Online Data Table 5” in **online\_data\_tables.xls.**

We computed the Latitude/Longitude for each CZ (shown in Columns D and E of **Chetty\_Data\_1.xlsx**) using two additional data sources:

* **czlma903.xls** contains the FIPS for counties in each of the CZs. This file is available at: <https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/>
* **county\_centers.csv** contains the latitude and longitude centers for each county (FIPS). This file is available at: <https://www.btskinner.io/data/spatial-data-and-scripts/>

Data from these files were copied into tabs in the file **CZ\_FIPS\_LL\_Data.xlsx**.

The Matlab program **CS\_LatLon.m** computes the latitude and longitude values for each CZ, by averaging the values for each county within the CZ using county population weights available from the population values available in **czlma903.xls**. These values are then manually copied into new tabs in **CZ\_FIPS\_LL\_Data.xlsx** and are also copied into the **Chetty\_Data\_1.xlsx** file.

**Chetty\_Data\_contiguousUS.csv** is a csv copy of **Chetty\_Data\_1.xlsx** with the rows relating to AK (Alaska) and HI (Hawaii) deleted.

Finally, **CZadj\_matrix.csv** contains the adjacency matrix of the 722 commuting zones in contiguous U.S.

* CZadj\_matrix.csv is computed using files in the folder **Data\AdjacencyMatrix\_cz1990\**. The details/sources are as follows:
  + The files **cz1990\_shapefile** and **cty\_cz\_st\_crosswalk.csv** were downloaded from https://healthinequality.org/data/
  + Run the R program **cz1990.R** which creates the file **CZadj\_matrix.csv**. This was then copied into the **Data\** folder.

## Computational requirements

### Software Requirements

* Stata (code was last run with version 18.0)
  + Figures 1 and 4 are produced in STATA using Michael Stepner's MAPTILE program. Information is available here: https://michaelstepner.com/maptile/
* Matlab (code was last run with Matlab Release 2024a)
  + Additional requirements: The programs use the Matlab function **distance**. This is available in the Mapping Toolbox, which should be installed.
* Intel Fortran Compiler version: ifort 19.1.3.311
  + Additional requirements: IMSL Numerical Libraries Version 7.01
* Mathematica (version: 13.0)
* R with “tidyverse”, “sf”, and “furrr” libraries installed. (See comments on the top of **cz1990.r** for information about downloading these libraries if they are not already installed on your computer.)

### Memory and Runtime Requirements

* See details below.

#### **Hardware**

* The Matlab and Stata code was last run on a 16-core Apple M3-Max with MacOS version 14.4.1 using a MacBook Pro with 128GB of memory.
* The Fortran and Mathematica code was run on a dual-12 core Windows 10 PC with 192 GB of memory.

## File structure

* All programs for organizing the Chetty et al data set are in the folder: **Data**.
* All programs used for Figures 1 and 4 are in folder: **STATA\_Matlab\_Examples**.
* All programs for constructing the analysis for Table 1 are in the folder: **Matlab**. This folder also contains the program used format the output for Table 2.
* The Mathematica notebooks used in the construction of Figures 2 and 3 and in the folder: **Mathematica**.
* The Fortran programs used to compute the simulated data reported in Table 2 are in the folder: **Fortran**.
* The simulation results summarized in Table 2 are stored in the folder: **mc\_files**.
* The folders **cscpc** and **matlab\_functions** contain Matlab scripts used by other programs.

## Instructions to Replicators

* Figure 1
  + Step 1: Run the Matlab script: **STATA\_Matlab\_Example\** **Figure1\_Figure4\_Examples\_Data.m** (Run time less than 1 minute)
  + Step 2: Run the Stata file: **STATA\_Matlab\_Example\Figure\_1\_4.do** (Run time less than 1 minute).
    - Note: the do file uses Michael Stepner's MAPTILE program. Information is available here: https://michaelstepner.com/maptile/
* Figures 2 and 3
  + Step 1: Evaluate the entire Mathematica notebook **Figures2and3.nb**. Runtime is 4 minutes.
    - Results saved in folder define in “path” variable at beginning of notebook.
  + Step 2: The 4 generated pdf panels were manually superimposed with the outline of the U.S. in **US\_WhiteScreen.pdf**, also generated by the notebook. Note that **US\_WhiteScreen.pdf** contains a white frame around the transparent shape of the contiguous U.S., so it appears entirely white when viewed in a standard pdf reader.
* Table 1
  + Run the Matlab script: **Matlab\Table\_1.m** (approximate run time is 15 minutes)
    - Results are saved in folder: **Matlab\output\**
* Table 2
  + Step 1: Evaluate the entire Mathematica notebook **Project\_LatLong.nb**. Runtime less than a minute.
    - Set “path” variable at beginning of notebook to **data** subfolder of replication package.
  + Step 2: Compile and run the fortran program in the fortran subfolder (it is one program with many modules in different files; see the **readme.txt** file in the **fortran** subfolder for additional details and compiler options). Runtime about 1 hour.
    - In the rpmod.f90 file, set “folder\_data” variable to point to subfolder **data** and set “folder\_out” to point to subfolder **mc\_files** of replication package prior to compiling.
  + Step 3: Run the Matlab script: **Matlab\Table\_2.m** (run time less than 1 minute).
    - Results are saved in folder: **Matlab\output\**

## References

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014): “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States,” *Quarterly Journal of Economics*, 129, 1553–1623.