

Replication files for: **Spatial Unit Roots and Spurious Regression**  
by Ulrich Müller and Mark Watson

Revised: April 2024

**Data:**

The paper uses data from the Chetty, Hendren, Kline and Saez paper listed in the references. The data used in this paper are in the Excel File: **Chetty\_Data\_1.xlsx** (also available as a CSV file.) These data are in the Data folder. (Documentation for the data are given below.)

Note: Figures 1 and 4 are produced in STATA using Michael Stepner's MAPTILE program. Information is available here: <https://michaelstepner.com/maptile/>

**Instructions for reproducing tables and figures:**

Figures 1 and 4: These are produced using Matlab and STATA. Programs are in the STATA\_Matlab\_Examples folder.

Figure 1, Panel a:

Data are generated and plotted in **TS\_RW\_Generate.m**.

Figure 1, Panel b:

Data generated in **Levy\_CZ\_Generate.m** with data saved in **Levy\_CZ.xlsx** (Figure 1)

Data are plotted in **Figure 1: Levy\_Figure\_1.do**

Figure 1, Panel c and Figure 4:

Data are computed in from **Chetty\_Example.m** with Data saved in **Chetty\_Example.xlsx**

Data for Figure 1 panel c and Figure 4 are plotted in **Chetty\_Example\_Figure\_1\_4.do**

Figures 2 and 3: These were generated by the Mathematica notebook **Figures2and3.nb** and then manually superimposed with the outline of the U.S. generated by the **Plot\_UnitedStates.nb** in the Mathematica subfolder.

Table 1: All results computed in **Chetty\_Regressions\_1\_ComputeResults.m** which is in the matlab folder. The Table is formatted in **Chetty\_Table\_1\_FormatTable.m**, also in the matlab folder.

Table 2: All calculations underlying Table 2 are carried out using Fortran. See the readme.txt file in the Fortran folder. The input for the Fortran code are the locations of the 722 commuting zones and state indicators (for clustering). The locations are first projected into two dimensions using standard cartography, which is carried out in the Mathematica notebook

**Project\_LatLong.nb** which generates the file **Chetty\_Data\_proj.csv** from the file **Chetty\_Data\_contiguousUS.csv** (the latter being identical to **Chetty\_Data\_1.xlsx**, except that the commuting zones in Alaska and Hawaii are dropped). The projection is necessary, because some of the data generating processes are not guaranteed to generate positive definite matrices

when distances are computed using the great circle formula. In addition, the Fortran program reads in the adjacency matrix for the commuting zone stored in **CZadj\_matrix.csv**. The output of the Fortran program are csv files that are read and tabulated by the program **mc\_table\_2.m** in the matlab folder.

### Data Documentation:

The data are taken from the replication materials for the Chetty, Hendren, Kline and Saez paper listed in the references. The replication materials 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 copied these data into the file **Chetty\_Data\_1.xlsx**.

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.

#### In **Chetty\_Data\_1.xlsx**:

- The variables in columns M-AT (*FracBlack* through *IncGrth*) are from tab “Online Data Table 8”.
- The variables in columns G-L (*AM* through *RMMale*) are from tab “Online Data Table 5”

We computed the Latitude/Longitude for each CZ (shown in Columns D and E) 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** then 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 copied into new tabs in **CZ\_FIPS\_LL\_Data.xlsx** and also copied into the **Chetty\_Data\_1.xlsx** file.