Replication files for "Natural amenities, neighborhood dynamics, and persistence in the spatial distribution of income"

Sanghoon Lee Jeffrey Lin

February 28, 2017

1 Citation and references

Suggested citation:

> Sanghoon Lee and Jeffrey Lin. "Natural amenities, neighborhood dynamics, and persistence in the spatial distribution of income," *The Review of Economic Studies*, forthcoming.

Note that our database uses in its construction information from sources listed in the References section of this document. Please cite them appropriately. Refer to Appendix B for details about the construction of the data set and which sources were used for each variable.

2 Online appendix

The file "OnlineAppendix.pdf" contains Appendixes A, B, and C.

3 Replication files

This package contains files necessary to replicate the results reported in the paper. The Stata .do file "figures_and_tables.do" replicates all of the figures and tables reported in the paper. The Stata .do file automatically creates a directory structure in `results/` using the "mkdir" command. To run the script, you will need to edit the directory name and the section flags at the beginning of the file.

	File name	File type	Description
code/	figures_and_tables	Stata do	Reproduce Figures and Tables
data/	Lee_Lin_data	Stata 14 data	Consistent-boundary tract data, 1880–2010
map/	figure1_color	ArcGIS 10 map	Reproduce Figure 1 in color
map/dat/	figure1data	csv	Automatically generated by figures_and_tables.do
results/	*		Automatically generated by figures_and_tables.do
simulations/	Monte Carlo simulations.nb	Mathematica	Mathematica program to run Monte carlo simulations in online Appendix C.
tract_xwalk_files/	*	csv	Normalize historical census tract data to 2010 boundaries

Table 1: File names and descriptions

The exception is Figure 1, which can be replicated using the ArcGIS map document "figure1_color.mxd." The map document references two files which are publicly available and not included in this package. The 2010 census tract shapefile can be obtained from NHGIS. The base map raster file can be obtained from the U.S.G.S. (2013).

4 Replication of simulation results

The Mathematica 11 script "simulations/Monte Carlo simulations.nb" replicates the simulation results reported in Appendix C. To run the program, open the program in Mathematica, select all cells by pressing 'Ctrl-A', and evaluate the selected cells by pressing 'Shift-Enter'. The accompanying pdf file "Monte Carlo simulations.pdf" shows a sample output where the Tables in Appendix C are generated in Out[14], Out[15], and Out[19].

5 Tract boundary normalization

A full description of our tract boundary normalization procedure can be found in Online Appendix B.

Census tract data from 1910–2010 can be accessed through NHGIS at `https://www.nhgis.org`. The NHGIS data selection tool allows filtering by geography—in this case census tract—and year. Census microdata from 1880 can be accessed through IPUMS USA at `https://usa.ipums.org/usa/`. These individual data can then be aggregated using the enumeration district identifiers.

To normalize 1970–2000 data to 2010 census tract boundaries, we use the Longitudinal Tract Database (LTDB) (Logan, Xu, and Stults, 2014). The LTDB can be accessed at `https://s4.ad.brown.edu/Projects/Diversity/Researcher/Bridging.htm|`.

To normalize 1880 and 1910–1960 data to 2010 census tract boundaries, we use weights determined by overlapping land area. Boundary files for 1910–1960 and 2010 census tracts can be accessed through NHGIS at `https://www.nhgis.org/documentation/gis-data`. Boundary files for 1880 census tracts can be accessed

through the Urban Transition Historical GIS (UTHGIS) project at `https://s4.ad.brown.edu/Projects/UTP/index.htm`.

These weights can be computed using files in the "tract_xwalk_files" directory, which contains 7 crosswalk files in CSV format, one for bridging each historical census year to 2010. Each crosswalk file contains 3 columns: (1) a 2010 census tract identifier, (2) a historical census tract or enumeration district identifier, (3) land area of overlap between the 2010 census tract and the historical census tract. These files were created using ArcGIS's "Intersect" tool, using the boundary files available from UTHGIS and NHGIS.

To normalize historical census tract or enumeration district data to 2010 census tract boundaries, we perform the following steps.

1. Create an extract from the NHGIS or IPUMS containing the desired variables, by historical census tract for 1910–1960 or aggregated to the enumeration district for 1880.

2. Perform a one-to-many merge between the NHGIS or IPUMS extract and the crosswalk file from that year using the historical census tract or enumeration district identifier. For example, in Stata:

```
. insheet using "tract_xwalk_files/1910.csv" /* crosswalk file */
. sort gisjoin
. merge m:1 gisjoin1910 using "data1910", k(3) /* from NHGIS */
```

3. Compute weights. In Lee and Lin (2017), we compute weights c_h as follows. Define $h \in H$ as elements of the set of non-overlapping polygons that result from intersecting census tract boundaries in historical year t and in year 2010. (Each polygon h is a row in the crosswalk file.) The land area of each polygon is a_h. The land area of each historical census tract f is partitioned into one or more 2010 tracts, with shares $c_h \equiv \frac{a_h}{\sum_{h \in f} a_h}$. In Stata:

```
. bysort gisjoin1910:  egen tractarea1910 = sum(area)
. gen ch1910 = area/tractarea1910 /* weight c_h */
```

4. For variables reported as totals (e.g., total population), multiply historical variable values by the weight c_h. In Stata:

```
. gen wtpop = totpop * ch1910
```

5. For variables reported as means (e.g., average household income), multiply historical variable values by the weighted historical total population of polygon h.

6. Sum (or average) the weighted historical variable values to obtain the values for 2010 tract boundaries. In Stata:

```
. bysort gisjoin2010:  egen totpop1910 = total(wtpop), missing
. egen tag = tag(gisjoin2010)
. keep if tag
```

6 References

Bogue, Donald. *Census Tract Data, 1940: Elizabeth Mullen Bogue File* [Computer file]. ICPSR version. University of Chicago, Community and Family Study Center [producer], 1975. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor] (2000a).

Bogue, Donald. *Census Tract Data, 1950: Elizabeth Mullen Bogue File* [Computer file]. ICPSR version. University of Chicago, Community and Family Study Center [producer], 1975. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor] (2000b).

Bogue, Donald. *Census Tract Data, 1960: Elizabeth Mullen Bogue File* [Computer file]. ICPSR version. University of Chicago, Community and Family Study Center [producer], 1975. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor] (2000c).

Federal Emergency Management Agency. *National Flood Hazard Layer* [Computer file] (2012). Data accessed on July 17, 2012.

Fee, Kyle, and Daniel Hartley. "The Relationship Between City Center Density and Urban Growth or Decline," in S. Wachter and K. Zeuli, eds., *Revitalizing American Cities*. Philadelphia: University of Pennsylvania Press (2013), 45–64.

Gyourko, Joseph, Albert Saiz, and Anita Summers. "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index," *Urban Studies*, 45:3 (2008), 693–729.

Kneebone, Elizabeth. *Job Sprawl Revisited: The Changing Geography of Metropolitan Employment.* Washington, DC: Metropolitan Policy Program at Brookings (2009).

Logan, John R., Jason Jindrich, Hyoungjin Shin, and Weiwei Zhang. "Mapping America in 1880: The Urban Transition Historical GIS Project," *Historical Methods*, 44:1 (2011), 49–60.

Logan, John R., Zengwang Xu, and Brian Stults. "Interpolating U.S. Decennial Census Tract Data from as Early as 1970 to 2010: A Longitudinal Tract Database," *Professional Geographer*, 66:3 (2014) 412–420.

National Geospatial-Intelligence Agency. *World Port Index, Twenty-third Edition*, Pub. 150 [computer file] (2014). Springfield, VA: National Geospatial-Intelligence Agency [producer and distributor].

National Oceanic and Atmospheric Administration. *Coastal Geospatial Data Project* [Computer file] (2012). `http://coastalgeospatial.noaa.gov/data_gis.html`. Data accessed on September 26, 2012.

Office of Management and Budget. *Metropolitan and Micropolitan Statistical Areas and Components, December 2009, with Codes* [Computer file] (2009). `http://www.census.gov/population/metro/files/lists/2009/List1.txt`.

Minnesota Population Center. *National Historical Geographic Information System: Version 2.0.* Minneapolis, MN: University of Minnesota, 2011.

PRISM Climate Group, Oregon State University. *PRISM Climate Data* [Computer file] (2004). `http://prism.oregonstate.edu`.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. *Integrated Public Use Microdata Series: Version 5.0* [Computer file]. Minneapolis: University of Minnesota (2010).

U.S. Geological Survey. *100-Meter Resolution Natural Earth of the Conterminous United States* [computer file]. Rolla, MO: National Atlas of the United States (2013).