

Replication Documentation for Glaeser and Gottlieb (2008)*

Edward L. Glaeser
Harvard University and NBER
eglaeser@harvard.edu

and

Joshua D. Gottlieb
Harvard University
jdgottl@fas.harvard.edu

August 26, 2008

1 Data Preparation

This section describes the acquisition and preparation of various data files for use in creating the tables and figures in Glaeser and Gottlieb (2008). File locations are described as on a Unix computer system, relative to a home directory `~/`. Various do-files documented in section 2 also refer to other files relative to the same paths, so researchers replicating any of the analysis will want to set up a similar directory structure. Our code, along with some of the raw data files, is available online in the appropriate location relative to the home directory at <http://www.nber.org/~jdgottl/BPEA/>.

The original data for this paper are primarily from the U.S. Census, in two forms. First, aggregate data have been released by the Census in various forms over the decades, including the City and County Data Books. We use the compilation of these releases by Michael R. Haines (2005), available from the Inter-university Consortium for Political and Social Research¹ as Study No. 2896. All of the files in this ICPSR study should be downloaded and placed in the folder `~/urban/census/icpsr/`.

*“The Economics of Place-Making Policies,” *Brookings Papers on Economic Activity*, Spring 2008

¹The ICPSR website is <http://www.icpsr.umich.edu/>.

The second major data source is the Census Public Use Microdata Sample. We downloaded these data from the Integrated Public Use Microdata Series (IPUMS)² of the University of Minnesota's Minnesota Population Center. The file

`~/bulk/ipums/jdgottl_fas_harvard_edu_010.cbk` describes the download from the IPUMS website. The downloaded file should be placed in `~/bulk/ipums/` and unzipped; then `~/bulk/ipums/jdgottl_fas_harvard_edu_010.do` should be run in Stata to convert the download into a Stata `.dta` file.

Other data sources include the files `~/urban/geography/zip/pcwi5.dta` from Glaeser and Kahn (2004), `~/urban/census/pm10.dta` from Kahn (2003)³, and `~/urban/census/murder.dta` downloaded from the FBI. Cost of living data from the American Chamber of Commerce Research Association (ACCRA) are to be mapped to 1999 MSA codes and placed in `~/urban/geography/accra data.dta`.⁴ Geography data are from Rappaport and Sachs (2003)⁵ and are saved in `~/urban/geography/coast_variables_01b.dta`, and presence in the Appalachian Regional Commission was inputted directly from the ARC website⁶ into `~/urban/geography/arc.dta`. The Craig, Palmquist and Weiss (1998) transportation data were provided by Robert Margo in `~/urban/transportation/Craig-Transport-Data.xls` and exported to `~/urban/transportation/Craig-Transport-Data.dta`. Nathaniel Baum-Snow provided the replication CD for Baum-Snow (2007); the relevant data are copied from his file `hwy-allyr-cnty.dta` into `~/urban/transportation/BaumSnowMSAhighways.dta`, and his file `county-change-code.do` is copied into `~/urban/geography/`. Federal transportation spending is in `~/urban/transportation/Transport_Aid_Revised.dta` and airport boardings are saved in `~/urban/transportation/msaairports.dta` from the Bureau

²The IPUMS website is <http://usa.ipums.org/>.

³These data are originally from http://www.epa.gov/aqspubli/annual_summary.html.

⁴These data were purchased from the ACCRA website at <http://www.coli.org/> and are private, so we only include the empty file `accradata(empty)` which illustrates the organization of the file.

⁵They can be downloaded from Jordan Rappaport's website at <http://www.kc.frb.org/econres/staff/jmr.htm>.

⁶<http://www.arc.gov/index.do?nodeId=27>

of Transportation Statistics⁷ Gyourko, Saiz and Summers's (2008) Wharton Land Use Index is available at Joe Gyourko's website⁸ and saved in `~/urban/cyclesdata/wharton2.dta`. Model Cities and Urban Renewal data were inputted directly from Staples (1970) into `~/urban/BPEA/MSAchars99pmsa` plus `urban_renewal.dta`. The file `~/urban/census/ToJoshToELG2.dta` containing happiness data by MSA comes from Glaeser, Gottlieb, and Redlick (2008, mimeo.)

2 Code Execution

2.1 Microdata Preparation

After unzipping the IPUMS download in `~/bulk/ipums/`, run `~/bulk/ipums/jdgottl_fas_harvard_edu_010.do` to import and label the data. (This file is mostly from the IPUMS automatic do-file generator, with some minor additions.) Then `~/urban/ipums/ipums.do` (which calls `defs99pmsa.do` to merge and renumber some MSAs) cleans and prepares the IPUMS data for micro-level regressions.

The first tasks to undertake with these microdata are the creation of certain industry-level and MSA/industry-level characteristics needed for Table 8, as well as MSA fixed effects for Figure 16. The first part is done in `~/urban/industries/msaindustries.do` and the MSA fixed effects are in `~/urban/industries/msaFEs.do`.

2.2 Aggregated Analysis

With the appropriate files in `~/urban/census/` and its various subdirectories, `~/urban/geography`, and `~/urban/transportation/`, the Stata do-file `~/urban/census/compilecensus.do` (which calls `~/urban/census/censusdata.do` and other files) creates a county-level dataset

⁷http://www.bts.gov/publications/national_transportation_statistics/html/table_01_41.html

⁸http://real.wharton.upenn.edu/~gyourko/WRLURI/WHARTON%20LAND%20REGULATION%20DATA_1_24_2008.zip

(`~/urban/census/ConsistentCountyData.dta`) and an MSA-level dataset (`~/urban/census/MSAData99pmsa.dta`; version with MSA fixed effects in `~/urban/census/MSAchars99pmsa.dta`) with various characteristics at these levels. By changing line 1279 of `censusdata.do` to use a different set of MSA definitions (the default is 1999 definitions, using PMSAs and NECMAs where appropriate; line 1280 or 1281 should be uncommented to use 2006 definitions at the MSA or Metropolitan Division level, respectively) an appropriate MSA data file (e.g., `MSAData06md.dta`) is created using different MSA definitions.⁹ We use the two different 2006 definitions for merging with other datasets using those definitions and creating some of the figures in the paper. The note to each figure lists which set of MSA definitions is used for that figure.

Once these files are created, the MSA-level analysis in `~/urban/BPEA/bpeatables.do` can be run to create Tables 1 and 4–8, as well as regressions reported in the text of the paper. The file `~/urban/BPEA/bpeagraphs.do` creates Figures 1–5, 7, and 10–16; Figure 6 is created separately in `Fig6.do` using GSS data and Figures 8 and 9 are directly from Glaeser and Kohlhase (2004). The graphs created in `bpeagraphs.do` use a program called `sfit.ado` which is provided in `~/urban/BPEA` and should be placed somewhere in Stata’s `ado-path`. Note that Figures 10 and 11 use data directly from Glaeser and Gyourko (2006), which we include as `~/urban/cyclesdata/Figure18data.dta`.

2.3 Microdata Analysis

The cross-sectional microdata regressions to create Tables 2 and 8 are run in `~/urban/ipums/ipumsxs.do` and the repeated panel regressions to create Table 3 are in `~/urban/ipums/ipumsts.do`. Both of these files call MSA data files created in section 2.2. The file `~/urban/ipums/vardecomp.do` decomposes the variance of individual income into variance across individuals and variance across MSAs, to generate a number cited in the text

⁹A similar file can be created for any MSA definitions by creating an appropriate file analogous to `~/urban/geography/countylist99pmsa.dta` and changing line 1279 to match the new file’s name.

of the paper.