Data Build for Problem Set 1

Owen Zidar

November 2018

This document describes the steps needed to replicate the dataset provided in Problem Set 1 of this portion of Graduate Public Finance.

1 Variables

1. **Population share**

   - Variable: popshare
   - Source: ACS 2016
   - Definition: share of total US population in each PUMA

2. **Employment**

   - Variable: emp
   - Source: ACS 2016
   - Definition: PUMA-level count of people aged 18-65, employed, not in the armed forces, who worked at least 40 weeks and an average of 35 hours per week last year, with positive income from wages

3. **Wages**

   - Variable: wage
   - Source: ACS 2016
   - Definition: average hourly wages by PUMA

4. **Rents**

   - Variable: rent
   - Source: ACS 2016
   - Definition: average monthly rent paid by PUMA

5. **Sales tax rate**

   - Variable: t_sales
   - Source: CSG Book of the State (1976-2017)
   - Definition: state-level general sales tax rate

6. **Corporate income tax rate**

   - Variable: t_corp
   - Source: CSG Book of the State (1976-2017)
   - Definition: state-level statutory corporate income tax rate

7. **Average effective personal income tax rate**

   - Variable: t_pinc
   - Source: ACS 2016, calculated using NBER TAXSIM (see Feenberg and Coutts, 1993)
   - Definition: the effective personal income tax rate is the ratio of state income tax liability to total income, averaged across wage-earners in each state.

8. **Measures of state amenities**

   - Variables: min_temp, max_temp, precip
   - Source: Couture, Duranton and Turner (2018)
   - Definition: these amenity variables are state-level minimum and maximum temperature, and average precipitation levels

2 Replication

2.1 Download ACS 2016 data

1. ACS data can be access on the IPUMS USA website. You must have an account to download a data extract.

2. Select the following variables:

Variable	Description
SERIAL	Serial number
HHWT	Household weight
STATEFIP	State FIPS code
PUMA	PUMA code
RENT	Monthly rent paid
PERNUM	Person number in the household
PERWT	Person sample weight
AGE	Individual age
EMPSTAT	Employment status
EMPSTATD	Employment status, detailed
WKSWORK2	Weeks worked last year
UHRSWORK	Usual hours worked last year
INCWAGE	Income from wages and salaries

2.2 Constructing the analysis dataset

The .do file *build_pset1.do* draws raw files from the *raw* folder in this replication file, and does the following:

2.2.1 Cleaning ACS 2016

1. Following Moretti (2011), restrict sample to individuals

   - Aged 18-65
   - Employed and not in the Armed Forces

- Who worked at least 40 weeks in the previous year, and who usually worked 35 hours per week
- With positive, non-missing wages
- With positive, non-missing rent

2. Collapse the ACS dataset to calculate average hourly wages, average monthly rent, population and employment levels by PUMA

2.2.2 Merge state tax variables

1. State sales and corporate income tax rates were digitized from the CSG Book of States. The files containing these rates are located in */raw/state_tax_rates/sales* and */raw/state_tax_rates/corp*, respectively.

2. Average effective income tax rates were calculated using the ACS 2016 dataset and NBER's TAXSIM calculator (see Feenberg and Coutts, 1993). See Section 2C of the build file *build_pset1.do* for the calculations of state income tax liability.

2.2.3 Merge amenities variables

1. The build file *build_pset1.do* selects some state-level amenities listed in Table A.5 of Fajgelbaum et al. (forthcoming), and sourced from Couture, Duranton and Turner (2018). The code keeps one observation per state for 2016, thus eliminating duplicates.

References

Couture, Victor, Gilles Duranton, and Matthew A Turner. 2018. "Speed." *Review of Economics and Statistics.* Forthcoming.

CSG Book of the State. 1976-2017. *Book of the States.* Council of State Governments.

Feenberg, Daniel, and Elisabeth Coutts. 1993. "An Introduction to the TAXSIM Model." *Journal of Policy Analysis and Management*, 12(1): 89–194.

Moretti, Enrico. 2011. "Local labor markets." In *Handbook of Labor Economics.* Vol. 4, 1237–1313. Elsevier.