Replication Package for "Bidding for Firms: Subsidy Competition in the U.S."

Cailin Slattery

December 4, 2024

1 Overview

This package contains the data and code for the replication of all of the tables and figures in "Bidding for Firms: Subsidy Competition in the U.S." The data come from a variety of sources, some publicly available, some collected by the author, and some that are publicly available given that the individual registers an account (this is the IPUMS USA Census data). For the IPUMS data, I provide the code and detailed instructions for the user to download the data, as well as the intermediate file created from the raw IPUMS data, but I do not provide the raw files.

In this document I provide detailed instructions for how to run the code and build the data for analysis, and then how to reproduce the tables and figures for the main text and the online appendix. The replication package also includes all the intermediate data files that are built from the raw data.

The data sources are cited in the Online Appendix.

2 Data Build and Analysis

In order to create the files needed for analysis you need to run `master_build.do`. In order to run the maximum likelihood estimation routine and perform the counterfactual analysis you need to run `Hv.R`. The intermediate data files are saved in the data folder (data/, which is suppressed when listing inputs and outputs below. The data subfolders raw/ and crosswalk/ house the raw data needed for analysis. The details follow.

Run `master_build.do`. This file sets the environment, downloads necessary packages (including one user written package: `adjust_inflation.do`), creates helper files, and then calls the following programs in order to build the data for analysis:

- process_ipums.do takes raw census data as an input, and then uses that data in order to create demographic and industry occupation variables at the county and county-naics level, respectively. This is commented out in the build, but can be uncommented once the reader has downloaded the raw data for themselves. There are two .do files that are called in process_ipums.do, these are make_demo.do and make_occ.do.
  - INPUTS: *raw census data*,
    crosswalk/'y'_2_fips.dta (created by author),
    raw/bls_emp_matrix_21.xlsx (BLS),
    crosswalk/2018-occupation-code-list-and-crosswalk.xlsx (Census),
    crosswalk/nem-occcode-acs-crosswalk.xlsx (BLS),
    crosswalk/ind_code_naics4_xwalk.xlsx (created by author)
  - *raw census data:* The following data set is extracted from https://usa.ipums.org/usa-action/variables/group
    - Sample: 2000 5% and 2001-2018 ACS
    - Variables: year, sample, serial, cbserial, hhwt, cluster, strata gq, statefip, countyfip, puma, metro, proptx99, pernum, perwt, race, raced, hispan, hispand, educ, educd, occ1990, occ2010, ind, ind1990, incwage
  - OUTPUTS: census_county_level.dta, naics4_occ_vars.dta
- build_cbp.do processes raw county business patterns data from the Census (https://www.census.gov/programs-surveys/cbp.html).
  - INPUTS: raw/cbp/cbp'y'co.txt [y to 2000 to 2016] (Census)
  - OUTPUTS: CBP_county_ind.dta, CBP_county_ind_n2.dta, CBP_county_ind_n3.dta
- make_industry.do creates wage and employment variables at the industry-county level.
  - INPUTS: raw/mega_threat_clean.xlsx (created by author), CBP_county_ind.dta, CBP_county_ind_n2.dta, CBP_county_ind_n3.dta
  - OUTPUTS: county_naics_rect.dta , wage_naics_rect.dta, cz_naics_rect.dta
- make_county.do assembles county and state level location characteristics from a variety of sources, to create a county-year-level data set of location characteristics.
  - INPUTS: raw/roadnetwork.txt (EPA Smart Location Database),
    raw/County_Transportation_Profiles.csv (BTS),
    raw/whartonlanduse2020.dta (Gyourko et. al 2019, https://real-faculty.wharton.upenn.edu/gyourko/land-use-survey/),
    raw/bea_county_stats.dta (personal income, population downloaded from BEA, https://www.bea.gov/data/by-place-us),

raw/county to state border.txt (Holmes 1998, `http://users.econ.umn.edu/~holmes/data/BorderData.html`),
raw/bls_county_unemp.dta (BLS, `https://www.bls.gov/lau/`),
raw/DEC_10_SF1_GCTPH1.US05PR_with_ann.csv (land area from Census Fact Finder),
crosswalk/county to cmsa crosswalk.xlsx (BLS),
crosswalk/cbsa2fipscty.csv (Census), raw/airports.xlsx (FAA),
university_rd.dta and raw/university_location.dta (NSF),
raw/taxrates_stateyear_1950_2017.dta (Slattery and Zidar 2020)

  - OUTPUTS: location_char.dta

- `make_subsidy.do` cleans the raw subsidy data hand-collected by the author, and adjusts subsidy amounts for potential non-discretionary incentives when applicable.

  - INPUTS: raw/state_level_spending.dta (created by author),
raw/bds_size_state.csv (Census Business Dynamics Statistics),
raw/mega_threat_clean.xlsx (created by author),
raw/state_raw_naics_RD.dta (Moretti and Wilson 2017, NSF),
raw/naics_multiplier.dta (Economic Policy Institute),
wage_naics_rect.dta
  - OUTPUTS: subsidy_data.dta

- `make_analysis.do` assembles data at subsidy, industry, and location level to create an analysis file.

  - INPUTS: cty2cz.dta, naics4_occ_vars.dta, location_char.dta, subsidy_data.dta, cz_naics_rect.dta
  - OUTPUTS: analysis_cz.dta, analysis_cz_long.dta

The resulting analysis files from the data build are **analysis_cz.dta** and **analysis_cz_long.dta**. These are the inputs into the main analysis of the paper and most of the supplementary figures and tables. Therefore, I will describe these files with a little bit more detail.

In **analysis_cz.dta** an observation is a location pair. One of the locations is always a subsidy deal winning location. Therefore, each observations has subsidy deal specific variables, the location of the subsidy deal winner, and a set of location characteristics (measured at the commuting zone level), with the suffix "_win". Each subsidy-deal winning location is paired with all possible alternative locations for the firm. So, each winning location will be observed approximately 700 times in the data set. Some of these possible alternative locations are the actual runner-up locations, as documented in the raw data. These are flagged. These possible alternative locations have the same set of location characteristics. So, for example, each observation will have a "corp_tax" variable, and a "corp_tax_win" variable, for the corporate tax in the potential location and the winning location respectively, in the year of the subsidy deal.

Lastly, there are a set of variables that take the difference between the potential and winning location. Here, we have, for example, "diff_corp_tax", defined as corp_tax - corp_tax_win.

In **analysis_cz_long.dta** there are no "diff" variables, and each observation is a location. In short, it is the "long" version of **analysis_cz.dta**.

Run `Hv.R`. This file performs the main analysis in the paper, and will produce the necessary input for all of the exhibits in Section 5 and 6 of the paper.

- INPUTS: data/analysis_cz.dta, data/analysis_cz_long.dta
- OUTPUTS: data/runner_up_correlations.csv, data/mle_coef_manuf.csv, data/sim_sub.csv, data/cf_locations.csv, data/runner_up_correlations_serv.csv, data/mle_coef_serv.csv, data/sim_sub_serv.csv, data/cf_locations_serv.csv
- FIGURES: draft/appendix/epdf_w.eps, draft/appendix/ecdf_w.eps, draft/appendix/epdf_v.eps, draft/appendix/ecdf_v.eps

3 Main Text Figures and Tables

The figures and tables are produced using `figures.do` and `tables.do`, respectively. I list the exhibits and the names of the input files for each figure and table:

Main Text Figures [`figures.do`]

1. Geographic Distribution of Subsidy-Giving and Incentive Spending
   n_sub_map.eps, cost_per_map.eps, spend_map.eps, spend_entry_map.eps
2. Subsidy Competition Example with 2 States
   no file, created in latex
3. Identification via Runner-up
   bs_subcorp.eps, bs_autonetwork.eps, bs_subunemp.eps, bs_subincome.eps
4. Model Fit: Runner-up Subsidy Offers and Predicted Valuations
   offer_v.eps
5. Runner-up $\hat{\pi}$ and $\hat{v}$ (Step 1)
   bs_manuf.eps, bs_serv.eps
6. Model Fit: Simulated Subsidy Competitions
   model_fit_whist.eps, modelfit.tex
7. Counterfactual Changes in Firm Locations
   map_real_manuf.eps, map_counter_manuf.eps, map_real_serv.eps, map_counter_serv.eps

Main Text Tables [`tables.do`]

1. Terms of Subsidy Deals
   ind_dstats.tex
2. Reduced Form Evidence: Determinants of Subsidy Size
   reg_reducedform.tex
3. Manufacturing Profit Parameters
   mle_coef_manuf.tex
4. Trade/Services Profit Parameters
   mle_coef_serv.tex
5. Welfare Analysis
   counter_v.tex

4 Online Appendix

In order to create all of the exhibits for the online appendix you need to run some additional analyses. These include `Hv_norc.R` (the model without random coefficients), `simulations.R` (the simulation of the model), and `Hv_altspec.R` (the alternate model as described in Appendix L). Then the remaining stata files create the tables and figures for the online appendix.

1. Run `Hv_norc.R` in order to create the input for Appendix Table K.4.
2. Run `tables_appendix.do` in order to create the majority of the tables in the appendix.
3. Run `figures_appendix.do` in order to create the majority of the figures in the appendix.
4. Run `simulations.R` in order to create the input for the figures in Appendix F (Figures F.1, F.2, F.3).
5. Run `figures_simulation.do` in order to create the figures in Appendix F (Figures F.1, F.2, F.3).
6. Run `Hv_altspec.R` in order to create the input for Appendix Section L (Figure L.2, Tables L.3, L.4, L.5).
7. Run `alternate_model.do` in order to create all of the figures and tables for Appendix Section L (Figures L.1, L.2, and Tables L.1, L.2, L.3, L.4, L.5).