Spatial Equilibrium, Search Frictions and Dynamic Efficiency in the Taxi Industry

README for replication

2021

Nick Buchholz
nbuchholz@princeton.edu

Programs used: Stata 16, Matlab R2020b, QGIS 2.5, shell scripts configured to run with the Slurm workload scheduler (https://www.schedmd.com) on the Della cluster at Princeton Research Computing. While most jobs can be run locally, the $SLURM_ARRAY_TASK_ID parameter enables parallelization of the workload. Replacing this input in the Matlab files with indexes referred by the "array" values will allow for checking code on local systems.

Note: Code uses root directory "/home/" to refer to folder containing replication files.

Data Availability Statement

The datasets used in this article are located in /home/data/taxi_trip_fare_0812.dta and /home/data/taxi_trip_fare_0912.dta

New York City Taxi and Limousine Commission 2012 Trip Data

The two monthly datasets included in the replication package (Buchholz 2021) have been obtained by the author directly from the New York Taxi and Limousine commission (TLC). Note: Comprehensive trip sheet data including origin and destination points and time stamps are available for direct download by the TLC (https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page).

2010 TIGER/Line Census Shapefiles

The replication dataset adds two additional fields containing location indices. The shapefile containing the GPS coordinates of these locations is included in /home/gis/Locations_39.shp and has been merged to pickup and drop-off coordinates in each TLC dataset using the Stata command gpsmap (Brophy 2013). The geographic shape files are drawn by the author by aggregating 2010 census tracts (see here for raw census tracts: https://catalog.data.gov/dataset/tiger-line-shapefile-2010-2010-state-new-york-2010-census-block-state-based-shapefile-with-hous).

GIS instructions

The GIS project file is located at /home/gis/nyc39.qgs. To reproduce mapped data: (1) import CSV output files as a layer into this project. (2) Copy Shapefile layer "Locations_39". (3) Join CSV data and copied Shapefile layer using id "Loc" (see [Layer Properties | Joins]). (4) Visualize data through [Layer Properties | Symbology] using graduated symbology.

Programs are to be run in the following order:

(1) Process raw data to prepare summary statistics and estimation

% (1.1) run programs/stata/1_export_main.do
Note: these program will produce the following CSV output. Data_Matches_08_2012.csv, Data_Matches_09_2012.csv, Data_OD_hour.csv, timefare.csv. Use Matlab to import and re-save each of these under the same names with suffix .mat for estimation. [Note: automation syntax for this procedure has changed across versions, so the manual procedure is recommended.]

% (1.2) Summary Statistics: run programs/stata/2_sumstats.do
% (1.3) Fare Effects: run programs/stata/3_fare_effects.do

(2) Analysis

% (2.1) Estimate sigma_epsilon: run "home/shell/est.sh"
- Performs grid search over sigma_epsilon using Matlab program estl_quick_bootstrap_rev4(sigepsin,x,month,bootstrap_yn). Input sigepsin is the estimate of interest, run for month={8,9} to get estimates for 08/2012 and 09/2012. Bootstrap parameter is set to bootstrap_yn=0. This shuts down the bootstrap resampling procedure when loading data.
- Estimate is the minimizer of criterion values
% (2.2) Estimate lambda and alpha for each month: run programs/ml/get_lambda.m
- Output file for use in demand estimation: LamLLT_2020_11500.csv
% (2.3) Estimate standard errors of model parameters: run "home/shell/est_bootstrap.sh"
- Run for month parameter set to 8 and 9
- Will run estl_quick_bootstrap_rev4.m with numerical optimizer to solve for sigma_epsilon (or "sigeps"). Bootstrap parmeter is set to bootstrap_yn=1 to enable resampling.

- Will also run estl_post_bootstrap.m to estimate vectors alpha and lambda for each sample and the associated std. deviation of all estimates to produce standard errors.
% (2.4) Demand Estimation: run programs/stata/2.4_demand_estimation.do
- Exports demand parameters with output "data/predict_xb_ijt_clean5.csv"

(3) Welfare results

% (3.1) Import predict_xb_ijt_clean5.csv and save as a matrix "data_fe" into file programs/ml/mat_data/data_fe_section.mat [manual process recommended.]
% (3.2) Convert key elements of the above output to matrix form: run programs/ml/create_fe_matrix.m
% (3.3) run programs/ml/compute_counterfactuals_eqm_btech_local.m
% (3.4) run programs/ml/make_snippets_results.m

(4) Compute counterfactuals

% (4.1) Initialize Pricing Counterfactuals run shell/cfw_X.sh for X = {1,2,3}
% (4.2) Technology Counterfactuals (MT): run programs/ml/compute_counterfactuals_eqm_btech_local.m with matching technology parameter set to 1 (mtech=1)
% (4.3) Technology counterfactuals (EI): run programs/ml/compute_counterfactuals_eqm_spp_local.m
% (4.4) Pricing counterfactuals: run programs/ml/cf_eqm_interpolation.m
- Outputs to .tex table inputs

Tables and Graphics

* Table 1 programs/stata/2_sumstats.do
* Table 2 programs/stata/2_sumstats.do
* Table 3 programs/stata/3_fare_effects.do
* Table 4 [notation summary]
* Table 5 programs/ml/output_bootstrap_results.m -> results_summary_table_2020.tex
* Table 6 programs/ml/direct_measures.m
* Table 7 programs/stata/2.4_demand_estimation.do [lines 147-159]
* Table 8 programs/ml/compute_counterfactuals_eqm_btech_local.m -> programs/ml/make_snippets_results.m
* Table 9 [descriptive exposition]
* Table 10 [Follow steps 4.2.1-4.2.4] -> make_snippets_cf.m

* Table A1 [descriptive] (within main tex file)
* Table A2 programs/stata/2_sumstats.do
* Table A3 programs/ml/consolidated_tables.m [section: Transition Matrix Tables]
* Table A4 /home/programs/stata/3_fare_effects.do
* Table A5 programs/ml/consolidated_tables.m [section: Alternate Initial Distributions]
  -comment line 13 and uncomment line 14 or 15 for each alternate initialization
* Table A6 shell/simvar.sh -> run /home/programs/ml/simvar_post.m
* Table A7 programs/ml/consolidated_tables.m [section: Transition Matrix Tables]
* Table A8 repeat step (4.1) and (4.4) with the following changes
  —change parameter "start_time" = 36 in /home/programs/ml/compute_counterfactuals_eqm_btech_local.m
  -change folders in cfw_X.sh to "robust"
  -change folders in cf_eqm_interpolation.m to "robust"
* Table A9 repeat step (4.1) and (4.4) with the following changes
  —change parameter "waiting_elasticity" to -0.8 in /home/programs/ml/compute_counterfactuals_eqm_btech_local.m
  -change folders in cfw_X.sh to "welas0p8"
  -change folders in cf_eqm_interpolation.m to "welas0p8"
* Table A10 repeat step (4.1) and (4.4) with the following changes
  —change parameter "waiting_elasticity" to -1.2 in /home/programs/ml/compute_counterfactuals_eqm_btech_local.m
  -change folders in cfw_X.sh to "welas0p8"
  -change folders in cf_eqm_interpolation.m to "welas0p8"

* Fig 1 /home/gis/nyc39.qgs
* Fig 2 /home/programs/ml/consolidated_plots.m
* Fig 3 [tikz plot within main tex file]
* Fig 4 /home/programs/ml/plot_matching_contours.m
* Fig 5 /home/programs/ml/consolidated_plots.m -> import doc/plots/eqSD.csv into QGIS
* Fig 6 /home/programs/ml/consolidated_plots.m
* Fig 7 /home/programs/ml/consolidated_plots.m
* Fig 8 programs/ml/direct_measures.m
* Fig 9 [tikz plot within main tex file]

* Fig A1 /home/gis/nyc39.qgs
* Fig A2 (stata, descriptive)

* Fig A3 (stata, descriptive)
* Fig A4 programs/ml/output_bootstrap_results.m
* Fig A5 /home/shell/simvar.sh -> run /home/programs/ml/simvar_post.m
* Fig A6 /home/programs/ml/consolidated_plots.m
* Fig A8 /home/programs/ml/consolidated_plots.m
* Fig A9 /home/programs/ml/consolidated_plots.m
* Fig A9 /home/programs/ml/consolidated_plots.m
* Fig A10 /home/programs/ml/consolidated_plots.m
* Fig A11 Steps (4.1)-(4.3) generate map data for each counterfactual. Repeat for each counterfactual price of interest as determined in step (4.4), inputting these into programs/ml/compute_counterfactuals_eqm_btech_local.m and save using commented lines 279-321. After this, run /home/programs/ml/export_map.data -> import to QGIS. See /home/gis/nyc39.qgs
* Fig A12 See Fig A11 instructions.

References

Buchholz, Nicholas. 2021. "Data for: Spatial Equilibrium, Search Frictions and Dynamic Efficiency in the Taxi Industry." Review of Economic Studies. https://doi.org/10.5281/zenodo.4907192

Brophy, T. (2013): "gpsmap: Routine for verifying and returning the attribute table of given decimal degree GPS coordinates," in 2013 Stata Conference, Stata Users Group, 14.