

Replication Instructions for A Dynamic Model of Demand for Houses and Neighborhoods

Patrick Bayer, Robert McMillan, Alvin Murphy, Christopher Timmins

(1) Data

The paper uses data drawn from a number of different sources to estimate the model. The primary household-level sample is based on a merge of housing transactions data from DataQuick (now owned by CoreLogic) with mortgage application information from the Home Mortgage Disclosure Act. Neighborhood amenity data are based on measures of (i) violent crime from the RAND California database, (ii) ozone levels from the California Air Resources Board, and (iii) racial composition from the Census Summary Files. The paper provides a detailed description of these data sources.

While some of the underlying data sources are proprietary, all of the final data files used for estimation are provided in the folder Est_Data.¹

As discussed in the paper, the measure of neighborhood used is based on an aggregation of Census tracts into larger regions of roughly equal-sized population. The following four files contain measures of amenities and aggregate housing prices for each neighborhood.

- i. neigh_race.raw. This file provides a measure of percentage white for every neighborhood in every year.
- ii. neigh_price_index.raw. This file provides a measure of a repeats-sales price index for every neighborhood in every year.
- iii. neigh_pollution.txt. This file provides a measure of pollution (ozone exceedances) for every neighborhood in every year.
- iv. neigh_crime.txt. This file provides a measure of violent crime for every neighborhood in every year.

Household-level data are contained in a single file:

- v. zdata.raw. The variables included are: year of purchase, race, income, down payment, inflation-adjusted purchase price, neighborhood, year of future sale, year of future purchase, and neighborhood of future purchase.

Finally, a portion of the analysis requires the mapping of neighborhoods to counties within the San Francisco Bay Area.

- vi. neigh_county_map.txt. This file contains data mapping neighborhoods to counties.

¹ The housing transactions data and crime data were purchased under license from DataQuick and RAND, respectively. The final data files used in estimation require only processed or aggregate variables derived from these original sources and, as a result, the data files provided here do not directly reveal any individual observations from these original data sources.

(2) Estimation

The Matlab programs needed to replicate the empirical results in the paper are also included in the folder Est_Data. Fully replicating the results requires running only a single program, master.m, which calls a number of the sub-files included in the folder.

To replicate the complete set of findings in the paper, run master.m in Matlab. (We used version 8.1.0.604 (R2013a)). This master file calls the relevant files for the estimation of the model with and without unobserved heterogeneity. To run the model with liquidity constraints, simply uncomment line 23 (S4est_liqcon). The master file also runs the bootstrap. Finally, it will output some results to the screen and compile the tables (in tex) that are included in the paper.

We briefly outline the key steps and sub-files here:

- i. master.m calls S12prep.m and S12est.m, which estimate the portion of the model relating to the household location and mobility decisions.
- ii. master.m calls S3prep and S3est, which estimate the transition probabilities, simulate the continuation value, and recover the flow utility estimates.
- iii. master.m calls S4prep and S4est, which decompose the flow utility function.

Further documentation is provided as comments within both master.m and its sub-files.