README

The code in this replication package reproduces the data used in the paper using Excel, Fortran, and Matlab. Part of the data comes from publicly available surveys not generated by the authors, and part of it is generated by simulating the quantitative model described in the paper. This README explains how to construct both. The compilation of the survey data is straightforward, and the only calculations involved are summary statistics. To generate the model data, the replicator should expect the code to run for about 40 hours.

**Data**

We certify that the authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.

We certify that the authors of the manuscript have documented permission to redistribute/publish the data contained within this replication package.

1. Survey data from external sources, not generated by the authors.

* The Data Appendix details how to access the data and describes the calculations. The *rent\_to\_income\_analysis.xlsx* Excel file compiles and summarizes the affordable housing data described in the Data Appendix.
* All data are publicly available. The main data sources follow:
* US Office of Management and Budget (metropolitan statistical areas delineation).
* US Census Bureau (American Community Survey, New York City Housing and Vacancy Survey).
* Zillow (house price and rental indices).
* Affordable Housing Online (county-level number of rent-stabilized units).

1. Data used for model calibration. The data used to run the code have been deposited in the *affordable\_migrU40\_base\_nodat* repository for the baseline model with migrations in the main text and in the *affordable\_nomigr\_base\_nodat* repository for the no migrations model in appendix. They are in *txt* format. They include both exogenous objects (e.g., exogenously calibrated parameters such as age-dependent mortality rates) and endogenous objects (e.g., coefficients of the forecasting rules to compute the equilibrium of the model as a fixed point between households’ forecasts and their realizations). The endogenous objects are produced by the model as described later in this README. The exogenous objects are taken directly from external data sources, which are described now.

* All data are publicly available. Details on data sources and calculations are in the main text and the Data Appendix. The main data sources follow:
* Census Bureau (survival probabilities from mortality rates).
* Survey of Consumer Finances 1983-2010 (labor income by age).
* IPUMS Census (labor income by productivity).
* Bureau of Economic Analysis fixed asset tables 1972-2016 (housing depreciation).
* Internal Revenue Service SOI Tax Stats Migration data 2011-2012 (migration rates), which is compiled in Excel file *Migrationbyageandincome.xlsx*.
* Referenced papers from which we directly take estimated parameters: Saiz (2010, housing supply elasticity); Heathcote, Storesletten, Violante (2017, progressive labor income tax); Davis and Ortalo-Magne (2011, Cobb-Douglas preference for housing).

Dataset list

|  |  |  |  |
| --- | --- | --- | --- |
| **Data file** | **Source** | **Notes** | **Provided** |
| rent\_to\_income\_analysis.xlsx | US Office of Management and Budget, US Census Bureau, Zillow, Affordable Housing Online | Combines multiple data sources, serves as input for model calibration, Table 1, Figure 4 | Yes |
| Migrationbyageandincome.xlsx | Internal Revenue Service SOI Tax Stats Migration | Serves as input for model calibration, Figure 1 | Yes |

Exhaustive data citations

These citations refer to data directly used in our computations (citations for data used as a motivation are in the main text).

Affordable Housing Online, 2002-2022. County-level data on the number of rent stabilized units. <https://affordablehousingonline.com>. Accessed March 21, 2022.

Albert Saiz. Replication data for: The Geographic Determinants of Housing Supply. The Quarterly Journal of Economics. August 2010. Volume 125, Issue 3, Pages 1253–1296, https://real-faculty.wharton.upenn.edu/gyourko/land-use-survey/. Accessed March 21, 2022.

Brookings Institute. [Benjamin H. Harris](https://www.urban.org/author/benjamin-h-harris) and [Brian David Moore](https://www.urban.org/author/brian-david-moore). November 18, 2013. Residential Property Taxes in the United States map. <https://www.brookings.edu/interactives/map-property-taxes-in-your-county/>. Accessed March 21, 2022.

[Board of Governors of the Federal Reserve System (U.S.)](https://www.worldcat.org/search?q=au%3ABoard+of+Governors+of+the+Federal+Reserve+System+%28U.S.%29&qt=hot_author). Survey of Consumer Finances, 1983-2010. <https://www.federalreserve.gov/econres/scf_2010.htm>. Accessed March 21, 2022.

Charlynn Burd, Michael Burrows, and Brian McKenzie. Travel Time to Work in the United States: 2019. March 18, 2021. *Report Number ACS 47*. Source: U.S. Census Bureau, 2006 and 2019 American Community Surveys, 1-year estimates. Accessed March 21, 2022.

Hyunseung Oh and Chamna Yoon. Replication data for: Time to build and the real-options channel of residential investment. *Journal of Financial Economics*. January 2020. Volume 135, Issue 1, Pages 255-269. <https://www.sciencedirect.com/science/article/abs/pii/S0304405X19301722>. Accessed March 21, 2022.

Internal Revenue Service. Statistics of Income Division (SOI) Tax Stats – Migration Data, 2011-2012. Gross Migration file. <https://www.irs.gov/statistics/soi-tax-stats-migration-data-2011-2012>. Accessed March 21, 2022.

[Laurie Goodman](https://www.urban.org/author/laurie-goodman), [Alanna McCargo](https://www.urban.org/author/alanna-mccargo), [Ellen Seidman](https://www.urban.org/author/ellen-seidman), [Jim Parrott](https://www.urban.org/author/jim-parrott), [Sheryl Pardo](https://www.urban.org/author/sheryl-pardo), [Jun Zhu](https://www.urban.org/author/jun-zhu), [Bing Bai](https://www.urban.org/author/bing-bai), [Karan Kaul](https://www.urban.org/author/karan-kaul), [Maia Woluchem](https://www.urban.org/author/maia-woluchem), [Bhargavi Ganesh](https://www.urban.org/author/bhargavi-ganesh), [Alison Rincon](https://www.urban.org/author/alison-rincon). *Housing Finance at a Glance: A Monthly Chartbook, October 2016*. The Urban Institute. 2016. Accessed March 21, 2022.

National Center for Health Statistics, National Vital Statistics System, Mortality.

Rong Wei, Ph.D., Office of Research and Methodology; Robert N. Anderson, Ph.D., Division of Vital Statistics; Lester R. Curtin, Ph.D., Division of Health and Nutrition Examination Surveys; and Elizabeth Arias, Ph.D., Division of Vital Statistics. September 2012. U.S. Decennial Life Tables for 1999–2001: State Life Tables. Volume 60, Number 9. <https://www.cdc.gov/nchs/data/nvsr/nvsr60/nvsr60_09.pdf>. Accessed March 21, 2022.

Steven Ruggles, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler and Matthew Sobek. IPUMS USA: Version 11.0 [dataset]. Minneapolis, MN: IPUMS, 2021. <https://doi.org/10.18128/D010.V11.0>. Accessed March 21, 2022.

U.S. Bureau of Economic Analysis. Standard Fixed Assets Tables, 1972-2016. <https://apps.bea.gov/iTable/index_FA.cfm>. Accessed March 21, 2022.

U.S. Census Bureau. American Community Survey 2015, 1-Year Data. <https://www.census.gov/data/developers/data-sets/acs-1year.2015.html>. 2006 to 2018, 5-Year Data. <https://www.census.gov/data/developers/data-sets/acs-5year.2018.html>. Accessed March 21, 2022.

U.S. Census Bureau. American Community Survey 2011-2015 ACS 5-Year Public Use Microdata Sample (PUMS). <https://data.census.gov/mdat/#/search?ds=ACSPUMS5Y2015>. Accessed March 21, 2022.

U.S. Census Bureau. U.S. Office of Management and Budget Delineation files. Last revised: October 8, 2021. <https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/delineation-files.html>. Accessed March 21, 2022.

U.S. Census Bureau. New York City Housing and Vacancy Survey (NYCHVS), 2017 Microdata. <https://www.census.gov/data/datasets/2017/demo/nychvs/microdata.html>. Accessed March 21, 2022.

Zillow. Housing Data. Zillow Home Value Index, Zillow Observed Rent Index. <https://www.zillow.com/research/data/>. Accessed March 21, 2022.

**Model generated data**

Most of the data used in the paper’s tables and figures is generated by simulating our quantitative model. These simulations are made after calibrating the model using survey data for the US and the NYC area, which was described above.

Note: to run the main Fortran code, a series of empty *dat* files must first be created, which will be populated with the age-dependent value and policy functions that the code solves for.

**Computational requirements**

Software:

* Matlab (the code was last run with Matlab Release 2021b)
* Fortran
* Intel Fortran compiler 19.1.2

The Fortran code was run on the NYU high-performance computing server (Greene). This requires shell scripting.

**Description of programs, files, and folders**

* The *run.sh* script runs the main Fortran code on the NYU high-performance computing server (Greene).
* The *affordable\_migrU40.f* code (in the *affordable\_migrU40\_base\_nodat* repository) solves for the steady state of the baseline with migrations. It also simulates a panel of agents. For each version of the model (baseline and policy experiments) the resulting simulated data is then downloaded locally to be processed by the Matlab codes, which are described below. Since it involves a large number of households and periods, these files are too large to be included here; they can be reproduced by running the code. These files should be deposited in the empty Results file in this replication package.
* The *affordable\_nomigr.f* code (in the *affordable\_nomigr\_base\_nodat* repository) does the same for the no migrations model in appendix.
* The *ModelOfCity20220105\_steadystate.m* code processes the simulation output created by the Fortran code. It creates the tables, figures, and numbers in the main text and the appendix.
* The *ModelOfCity20220105\_welfare.m* code computes the welfare changes associated with the policy experiments. It creates the welfare figures in the main text and the appendix.
* The *nwest.m* code computes a Newey-West adjusted heteroscedastic-serial consistent Least-squares regression, which is used as a check in the main Matlab code to compute the R-squared of forecasting rules.
* The *gini* folder contains a code which computes the Gini coefficient of a variable in a given population.
* The Figures folder contains all the figures, tables, and numbers in the main text and the appendix. They are created by the two Matlab programs described above.

**Which produced file belongs to which exhibit.** The code *ModelOfCity20220105\_steadystate.m* produces the tables and figures in the manuscript:

* Manuscript Figure 1 is migrOutMigRate.eps and migrInMigRate.eps
* Manuscript Figure 2 is migrvizbench\_v2o.eps
* Manuscript Figure 3 is migrcity\_fig\_wealth\_nyc\_MODEL.eps
* Manuscript Figure 4 is migrRCIncNWDist\_withdata.eps
* Manuscript Figure 6 is MigrRate\_barnew1\_NEWTABLE3.eps

***Table creation.*** For the baseline model and each policy experiment, the code produces a LaTeX file containing a set of moments, which are used in the main text and in Table 1 and Table 2. The resulting LaTeX files are deposited in the Figures file, and they are loaded at the beginning of the LaTeX file which produces the manuscript. These files are too numerous to be all listed. For instance, Table2migr.tex contains the moments for the baseline model with migrations, and Table3migrxpshareonefifty.tex contains the moments for the policy experiment with a share of rent-stabilized housing x 1.50. In each of these files, each moment receives a specific label. For instance, Table2migr.tex creates the label \NYCpopmigr for the value for the population of New York City generated by the simulation of the baseline model. These labels are then used in the main text and in Table 1 and Table 2, in the LaTeX file which produces the manuscript.

The code *ModelOfCity20220105\_welfare.m* produces the tables and figures in the manuscript:

* Manuscript Figure 5 is migrWelf\_main\_barnew1\_tr\_NEWTABLE3.eps

**Memory and runtime requirements**

The Fortran code was last run on the NYU high-performance computing server (Greene). We requested 48 CPUs and a memory allocation of 60GB. One iteration of the code (for a given set of coefficients for households’ forecasting rules) takes around two hours. The numbers of iterations required to solve for the full steady state of the model (including forecasting rules themselves) ranged between 20 and 40, depending on the policy experiment; they depend on the speed at which the coefficients are updated between two consecutive iterations (we used weights of 15% vs. 85% on the new vs. new coefficients to ensure smooth convergence). This process was repeated for each policy experiment and for the no migrations model, in addition to the baseline model.

The Matlab code was last run locally on a 2.3 GHz Quad-Core Intel Core i7MacBook Pro with macOS Big Sur.

**Instructions**

* On the high-performance computing server, create a series of empty dat files, which will be populated with the age-dependent value and policy functions that the code solves for.
* Run the run.sh script, which will run the affordable\_migrU40.f code (baseline model) on the server until the model converges. This produces simulated data in Matlab format for a large panel of households. This data can be replicated by running the code; it is too large to be included here.
* Download the simulated data to a local computer. Then, first, run the ModelOfCity20220105\_steadystate.m code to produce the figures, tables, and numbers in the paper; second, run the ModelOfCity20220105\_welfare.m code to produce the welfare figures.
* To compute the policy experiments, modify the corresponding parameters in the Fortran code and repeat the process. To create all the figures, tables, and numbers which involve a comparison between the baseline model and the policy experiments, the results from the baseline model must be kept in the local computer’s memory and not cleared.
* Repeat the same process for the no migrations model in appendix.