Replication Package Readme for "The 2000s Housing Boom With 2020 Hindsight: A Neo-Kindlebergerian View"

Gabriel Chodorow-Reich, Adam Guren, and Timothy McQuade

March 2023

This Readme document explains the code used in our paper. Section 1 provides our data availability statement. Section 2 contains our computational requirements. Section 3 explains the code for the results in sections 2 and 4 and Appendix B. Section 4 explains how we solve the model computationally and explains the files used to calculate the global model solution and impulse responses. Section 5 explains how we conduct our simulated method of moments estimation and calculate our final model results. Section 6 provides detailed file-by-file summaries for the model replication package.

1 Data Availability Statement

This section summarizes all of the data we use in the paper and its availability. We describe how we use each data set in subsequent sections.

- From Fred (Federal Reserve Bank of St. Louis, 2022), we obtain 4 series using Stata's "import fred" command:
  - Case-Shiller National House Price Index: Series CSUSHPINSA (accessed August 2021)
  - GDP Price Index: Series GDPCTPI (accessed August 2021)
  - GDPI Price Index: Series A191RG3A086NBEA (accessed November 2020)
  - Nominal Mortgage Rate: Series MORTGAGE30US (accessed September 2022)
- We obtain the Freddie Mac House Price Indices (Freddie Mac, 2021) for our main analysis at the CBSA level. These data can be directly downloaded from https://www.freddiemac.com/research/indices/house-price-index.
- We obtain the FHFA 5-Digit ZIP Code House Price Indices (Federal Housing Finance Agency, 2020) four our ZIP-code level analysis. We use the annual, not seasonally

adjusted "developmental" index, which can be downloaded at https://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_AT_BDL_ZIP5.xlsx.

- We obtain the CoreLogic House Price Indices (CoreLogic, 2018) for robustness at the CBSA level. This data is available for purchase from CoreLogic. Researchers interested in obtaining this data should contact the academic sales team; in 2021 we contacted Alec Henderson at alechenderson@corelogic.com. It can take several months to negotiate data use agreements and gain access to the data. The specific data series we uses is the Tier 11 Single Family Combined CBSA house price index as of September 2018.

- We obtain the ZIP-code level Zillow Home Value Index (Zillow, 2022). The data may be downloaded at https://www.zillow.com/research/data/. We use the all homes, seasonally adjusted series for our calculation of the downtown premium. We also obtain the CBSA-level Zillow Home Value Index for some robustness, which can be downloaded from the same source.

- We obtain survey measures of median inflation expectations form the Michigan Survey (University of Michigan Surveys of Consumers, 2022). We use Table 33, which can be downloaded from https://data.sca.isr.umich.edu/data-archive/mine.php.

- We use the 1995 Survey of Consumer Finances from 1995 (Federal Reserve Board of Governors, 1995) for the 1995 loan balance distribution. The data can be downloaded from the Federal Reserve at https://www.federalreserve.gov/econres/files/scfp1995s.zip.

- We obtain the Land Share from Larson et al., 2021. The data are available for download from the Federal Housing Finance Agency at https://www.fhfa.gov/PolicyProgramsResearch/Research/Pages/wp1901.aspx. We use the data as updated in November, 2020.

- We obtain measures of the number of housing units form the Census' annual estimates of housing units for counties in the United States United States Census Bureau, 2022. The 2010-2019 data we use can be found at https://www2.census.gov/programs-surveys/popest/tables/2010-2019/housing/totals/CO-EST2019-ANNHU.xlsx. The 2000-2010 data can be found at https://www2.census.gov/programs-surveys/popest/tables/2000-2010/intercensal/housing.

- We obtain the Wharton Residential Land Use Regulatory Index (Gyourko et al., 2008) from Wharton at http://real-faculty.wharton.upenn.edu/wp-content/uploads/~gyourko/WRLURI/WHARTON%20LAND%20REGULATION%20DATA_1_24_2008.zip.

- We compute population density from the United States Census by dividing county-level population estimates (United State Census Bureau, 2022) by land area (United States Census Bureau, 2011). The population counts can be downloaded from https://www.census.gov/programs-surveys/popest.html. The land mass is variable LND110200D and can be downloaded from https://www.census.gov/library/publications/2011/compendia/usa-counties-2011.html#LND.

- We use the County Business Patterns (United States Census Bureau, 2020) data to calculate our Bartik shocks for employment and wages. The CBP data can be downloaded at https://www.census.gov/programs-surveys/cbp/data.html. We also calculate restaurant density using the CBP data.

- We use weather data on January Temperature, January Sunlight, and July Humidity from the U.S. Department of Agriculture's Natural Amenities Scale (United States Department of Agriculture, 2022). The data can be downloaded from https://www.ers.usda.gov/data-products/natural-amenities-scale/.

- We use land unavailability from Lutz and Sand, 2019. This data can be downloaded from https://github.com/ChandlerLutz/LandUnavailabilityData.

- We use a secondary land unavailability measure from Saiz, 2010. This can be obtained from the author.

- We apportion 1990 Census population counts to 2010 tract definitions using the Longitudinal Tract Database (Logan et al., 2022). These data can be downloaded from https://s4.ad.brown.edu/Projects/Diversity/Researcher/Bridging.htm.

- We obtain several variables from IPUMS USA (Ruggles et al., 2022), available at https://usa.ipums.org. These include:
  - Mean rent by CBSA, where Census rent is the variable rentgrs and the mean is a weighted average by the variable hhwt.
  - The 1990 national college share by industry.
  - Actual employment growth by industry between the 1990 Census and 2019 ACS.

- We calculate the non-traditional Christian share in 1990 from data on church membership from the Association of Statisticians of American Religious Bodies (Statisticians of American Religious Bodies, 1992). The data can be downloaded at https://www.thearda.com/data-archive?fid=CMS90CNT.

- We use data on local government expenditures from the Census (United State Census Bureau, 1993) to create a ratio of inspection expenditures to tax revenue. The data is available for download at https://www.census.gov/data/datasets/1993/econ/local/public-use-datasets.html. We use codes E66, F66, G66, K66, L66, M66, N66, and R66 as inspection categories.

- We use data on the 1990 College share from the Longitudinal Tract Database (Logan et al., 2022). These data can be downloaded from https://s4.ad.brown.edu/Projects/Diversity/Researcher/Bridging.htm.

- 1990 national collage share by industry from IPUMS USA, predicted employment growth by industry from IPUMS.
- We use Home Mortgage Disclosure Act (HMDA) data (Consumer Financial Protection Bureau, 2022) to calculate the mean non-owner occupier share of purchase mortgages as in Gao et al., 2020. The HMDA data can be downloaded from the Consumer Financial Protection Bureau.
- We use proprietary deeds and assessor records from DataQuick to calculate to calculate the repeat sales standard deviation moment for our calibration. DataQuick no longer exists as a company, but the archive of the DataQuick data we use in our analysis is maintained by the Harvard Kenned School's Taubman Center for State and Local government and detailed in Online Appendix B of Guren, 2018.
- We use data on expectations from Shiller and Thompson, 2022, Table 3. We hand code the averages from their paper into our Matlab files.

2 Computational Requirements

Empirics

- Stata (code last run with version 17 on a desktop computer with 32 GB of RAM). The following packages are required:
  - estout (as of October 2022)
  - ivreg2 (as of October 2022)
  - reghdfe (as of October 2022)
  - xtbreak (as of October 2022)
  - Other ado files are included in folders.

Model

- All model files were run on the Boston University Shared Computing Cluster (SCC) using Matlab (code was run with release 2020b).
- We run our SMM in parallel because each combination of parameters takes several minutes and we run millions of combinations, so running them in parallel and then pooling the results allows things to run in a few days rather than many months. The final runs to make the figures used in the paper can be done on a standard desktop.
- The BU SCC uses a batch scheduler through the Sun Grid Engine. We include the shell files used to submit jobs to this scheduler in the model/shell_files folder. We typically submit 3,000 12-hour jobs for single core nodes with 16GB of RAM, as on the BU SCC free nodes are most available for short and simple jobs like this and we could typcially

obtain 1,000 nodes at a time, meaning the task ran in at most 36 hours. You can find details about how the SCC works to compare it to your high-performance computing cluster at https://www.bu.edu/tech/support/research/system-usage/.

3 Empirical Replication Package for Sections 2 and 4 and Appendix B

The subfolder "data/code" contains eponymous Stata .do files to create each table and figure in Sections 2 and 4 and Appendix B. For example, running fig1.do creates figure 1 of the paper. The file masterfile.do creates all of the tables and figures in these sections. The file masterfile.do also calls two files, smm-data.do and ltv-scf.do, that produce data sets used in the model estimation. The remainder of this section describes in detail the data used in each of these files. In many cases the first part of the code to reproduce the cleaned data set is turned off by default but provides additional line-by-line detail on how the data sets were constructed. The programs were run on Stata 17 and require the user to install `ivreg2`, `reghdfe`, and `estout`, available from `ssc`. Any non-standard `.ado` programs are included in the folder.

- `fig1.do`: The data for figure 1 are the Case-Shiller National House Price Index (FRED code CSUSHPINSA) and GDP price index (FRED code GDPCTPI).
- `fig2_figB4.do`: The data for figure 2 come from FHFA zip code house price indexes (`https://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_AT_BDL_ZIP5.xlsx`) and the GDPI price index (FRED code A191RG3A086NBEA). The data for figure B.4 come from the file `cbsa-data-cleaned.dta` described below.
- `fig3.do`: The data for figure 3 come from the BLS. The series IDs contained in `data/data/A_raw/BLS/cpi-rent-series_id.txt` were copied into `https://data.bls.gov/cgi-bin/srgate` to obtain the time series of CPI rent data by city. These data are adjusted using the suggested adjustments in Crone et al. (2010).
- `fig4_tab1_fig5.do`: The data for figures 4 and 5 and table 1 come from the file `cbsa-data-cleaned.dta` described below.
- `figB1_tabB1.do`: The data for figure B.1 and table B.1 come from the file `cbsa-data-cleaned.dta` described below.
- `tabB2.do`: Table B.2 reports results from tests for structural breaks in rent and price growth. The rent data were created in `fig3.do`. The price break data are created by running the file `cbsa-breaks.do` included in this replication package, which loads quarterly house price indexes from `https://www.freddiemac.com/research/indices/house-price-index`. These files require installing the Stata package `xtbreak` (`https://janditzen.github.io/xtbreak/`).

- tabB3_tabB4_tabB5_figB3_figB5.do: The data for tables B.3, B.4, and B.5 and figures B.3 and B.5 come from the file cbsa-data-cleaned.dta described below.

- figC3.do: The data in figure C.3 are the national house price series created in fig1.do, the nominal mortgage rate (FRED code MORTGAGE30US), and Michigan Survey median inflation expectations (table 33, downloaded from https://data.sca.isr.umich.edu/data-archive/mine.php).

- smm-data.do: This file processes data from the file cbsa-data-cleaned.dta described below and outputs a data set of moments for the SMM routine.

- ltv-scf.do: This file calculates the 1995 LTV distribution used to initialize the model. The SCF data are from https://www.federalreserve.gov/econres/files/scfp1995s.zip.

As noted above, our main analysis file used in fig4_tab1_fig5.do, figB1_tabB1.do, tabB3_tabB4_tabB5_figB3_figB5.do, and smm-data.do is cbsa-data-cleaned.dta. This file combines data from several sources. The files urbanization-data.do and cbsa-merged-data.do process the raw data files. The file prep-cbsa-data.do creates several derived variables used in the analysis and saves cbsa-data-cleaned.dta. We now explain the origin of each variable in the final data set that is used in the analysis, grouped by category.

Outcome and endogenous variables

- rhpifm_cbsa: Freddie-Mac house price index from https://www.freddiemac.com/research/indices/house-price-index.

- land_share2012: 2012 land share from https://www.sciencedirect.com/science/article/pii/S0304393220301379.

- units_1997_2019: Census housing units growth using units from https://www2.census.gov/programs-surveys/popest/tables/2010-2019/housing/totals/CO-EST2019-ANNHU xlsx and https://www2.census.gov/programs-surveys/popest/tables/2000-2010/intercensal/housing.

- z_WRLURI2006_pop: http://real-faculty.wharton.upenn.edu/wp-content/uploads/~gyourko/WRLURI/WHARTON%20LAND%20REGULATION%20DATA_1_24_2008.zip collapsed into CBSAs.

- Sl_downtownpremium_1997_2019: this variable is based on zip code house price data from FHFA already processed in fig2_figB4.do, from Zillow (https://www.zillow.com/research/data/), and from the Decennial Census and ACS, combined with assignment of zip codes to downtowns by ranking all Census tracts in the CBSA by their distance to the downtown center and defining the downtown as those tracts covering the closest 5% of population using 2010 tract definitions and 1990 Census population counts apportioned to 2010 tract definitions by the US2010 Project webpage (http://www.s4.brown.edu/us2010/Researcher/Bridging.htm). The downtown premiums for each data source are constructed in `urbanization-data.do` and combined into a single measure in `prep-cbsa-data.do`.

- `rent_census`: Mean Census rent by CBSA, where Census rent is the variable `rentgrs` in IPUMS and the mean is weighted by the variable `hhwt`.

Excluded instruments

- `totalunavland_totarea_pct`: Land unavailability from https://github.com/ChandlerLutz/LandUnavailabilityData.

- `popdensity`: Census population (https://www.census.gov/programs-surveys/popest.html) divided by land mass (variable `LND110200D` from https://www.census.gov/library/publications/2011/compendia/usa-counties-2011.html#LND).

- `CBPBartik4_cwage_F20` and `CBPBartik4_emp_F20`: We combine the County Business Pattern (CBP) files provided by the Census (https://www.census.gov/programs-surveys/cbp/data.html) with the files from Eckert et al. (2020) that optimally impute suppressed employment cells and provide a consistent correspondence to NAICS 2012. We use 1998 rather than 1997 as the initial year because the NAICS version of the data start in that year. The final year of data available is 2018. We implement "leave-one-out" shift shares: defining $E_{i,j,0}$ as employment in area i and industry j as a share of total date 0 employment in area i, $g_{-i,j}$ as the growth rate of employment in industry j in all other areas between dates 0 and 1, $w_{-i,j,t}$ as the wage (payroll per employee) in industry j in all other areas at date t, and $\hat{E}_{i,j,1} \equiv E_{i,j,0} \times g_{-i,j} / \left[\sum_k E_{i,k,0} \times g_{-i,k}\right]$ as the predicted date 1 area i employment share in industry j, the shift-share for the growth of employment is $\sum_j E_{i,j,0} g_{-i,j}$ and the shift-share for the growth of the average wage is $\left[\sum_j \hat{E}_{i,j,1} \times w_{-i,j,1}\right] / \left[\sum_j E_{i,j,0} \times w_{-i,j,0}\right] - 1$, where date 0 is 1998 and date 1 is 2018. For area-industries with suppressed wage data, we replace $w_{-i,j,t}$ with $w_{j,t}$, where $w_{j,t}$ is the national wage in industry j at date t.

- `JanTemp,JanSunlight,JulHum`: https://www.ers.usda.gov/data-products/natural-amenities

- `CBPempRestaurants_pc`: CBP restaurant density is NAICS codes 7725 or 7221 or SIC code 5812 divided by population.

- `share_nontrad`: Non-traditional Christian share from https://www.thearda.com/Archive/Files/Descriptions/CMS90CNT.asp, where traditional denominations are `CATHOL_A,UCC_A,AMBAPT_A,NABC_A,ARPC_A,CPC_A,EVPC_A,PCUSA_A,PCIA_A,AMEZ_A ,AWMC_A,EVMETH_A,FMETH_A,FUMETH_A,PRMETH_A,UNMETH_A,APOSLU_A,LUBRET_A, LUCONF_A,EELUTH_A,ELCA_A,ELSYN_A,FREELU_A,LATVLU_A,LCMS_A,AALUTH_A,PRCONF_A, WELS_A,REPISC_A,AMEZ_A,EPISC_A`.

- share_insp: inspection expenditures/tax revenue from https://www.census.gov/data/datasets/1993/econ/local/public-use-datasets.html, where inspection codes are E66,F66,G66,K66,L66,M66,N66,R66.
- colshareXrestaurants: 1990 college share × ratio of restaurant density in downtown to non-downtown, constructed in urbanization-data.do using Census college share from http://www.s4.brown.edu/us2010/Researcher/Bridging.htm and restaurant density from County Business Patterns.
- downtownXdcolshare: 1990 difference between [share of college-educated who live downtown] and [share of non college-educated who live downtown] × predicted change in college share, constructed in urbanization-data.do using Census college share from http://www.s4.brown.edu/us2010/Researcher/Bridging.htm and predicted change in college share based on 1990 national college share by industry from IPUMS, employment-by-industry in each CBSA from CBP, and predicted employment growth by industry using IPUMS.

Other variables

- hpicl_cbsa: CBSA house price indexes from CoreLogic. These data are proprietary and are replaced with the Freddie Mac data for purposes of this replication package. These data can be purchased from CoreLogic. Ask for the Tier 11 Single Family Combined House Price Index at the CBSA level. Our data were provided to us by CoreLogic on September 17, 2018. Import_hpi.do loads these from the raw file from CoreLogic.
- hpifhfa_cbsa: CBSA house price indexes from FHFA (https://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_AT_BDL_cbsa.xlsx).
- hpizhvi_metro:CBSA house price indexes from Zillow (https://www.zillow.com/research/data/).
- unaval: Land unavailability from Saiz (2010).
- jan2007equitysharelt: Equity distribution from Beraja et al. (2019) (received via private correspondence with authors and provided with permission).
- mean_gsx: Based on HMDA data (https://www.consumerfinance.gov/data-research/hmda/historic-data/).

4 Model Computational Methods

In this appendix, we detail the computational approach we employ for solving the model. We first discuss approximating the Fokker-Plank partial differential equation via polynomial projection methods. We then discuss how we use a sparse grids approach for our global solution method. Finally, we discuss how Monte-Carlo simulation methods are employed in our iterative global solution method to compute equilibrium mortgage pricing.

4.1 Approximating the Fokker-Plank Equation

The infinite-dimensional distribution of mortgage balances is a key state variable in our model and needs to be tracked along the equilibrium impulse response. The measure density $g(M,t)$ of this distribution follows the Fokker-Plank equation:

$$\frac{\partial}{\partial t} g(M,t) = (I_t + \iota) H_t \phi(M/P_t)/P_t - \iota g(M_t) \tag{1}$$

where ι is the arrival rate of liquidity shocks and $\phi(\cdot)$ is the LTV origination distribution. We approximate the loan balance distribution with a Chebyshev series:

$$g(M,t) = \sum_{i=0}^{N} \alpha_i(t) T_i(M),$$

$$\text{with } T_i(M) = \cos\left(i * \arccos\left(\frac{M - \left(M^u + M^l\right)/2}{\left(M^u - M^l\right)/2}\right)\right)$$

the (scaled) i^{th} Chebyshev polynomial of the 1st kind and M^u, M^l the upper and lower bounds of the mortgage balance distribution. The coefficients $\alpha_i(t)$ are set so that the polynomial series interpolates the true measure density at $(M_1, ..., M_N)$ collocation points. We set the collocation nodes to be the (scaled) Chebyshev-Gauss-Lobato (CGL) points:

$$M_i = \frac{1}{2}\left(M^u + M^l\right) - \frac{1}{2}\left(M^u - M^l\right)\cos\left(\frac{i\pi}{N}\right),$$

equal to the $(N-1)$ extrema of the N^{th} Chebyshev polynomial plus the endpoints. We thus require:

$$\sum_{i=1}^{N} \alpha_i(t) T_i(M_j) = g(M_j, t) \text{ for } j = 1, ..., N.$$

Using (1), we find the system of differential equations governing the coefficients:

$$\sum_{i=0}^{N} \alpha_i'(t) T_i(M_j) = (I_t + \iota) H_t \phi(M_j/P_t)/P_t - \sum_{i=0}^{N} \iota \alpha_i(t) T_i(M_j).$$

Letting $\mathcal{A}^* T_i(M) = \iota \alpha_i(t) T_i(M)$ we have:

$$\boldsymbol{\alpha}'(t) = T(\boldsymbol{M})^{-1}\left[\mathcal{A}^* T(\boldsymbol{M})\boldsymbol{\alpha}(t) + (I_t + \iota) H_t \phi(\boldsymbol{M}/P_t)/P_t\right],$$

$$\text{where: } \boldsymbol{\alpha}(t) = \begin{bmatrix} \alpha_0(t) \\ \vdots \\ \alpha_N(t) \end{bmatrix}, \; T(\boldsymbol{M}) = \begin{bmatrix} T_0(M_0) & \cdots & T_N(M_0) \\ \vdots & \ddots & \vdots \\ T_0(M_N) & \cdots & T_N(M_N) \end{bmatrix},$$

which provides a finite-dimensional system of differential equations governing the evolution of the coefficients of the Chebyshev expansion.

4.2 Mortgage Points Pricing and Sparse Grids

We use global solution methods to compute the equilibrium mortgage points. We then use this construction of mortgage points when calculating the equilibrium impulse responses which are the focus of our paper. Full grid methods for the global solution quickly run into the curse of dimensionality. We thus employ sparse grid techniques to get the global solution of the model. Our approach follows Judd et al. (2014).

We use the Smolyak construction for the sparse grids, once again utilizing Chebyshev-Gauss-Lobatto (CGL) points, that is extrema of Chebyshev polynomials of the 1st kind. In particlar, let d denote the number of state variables. The Smolyak construction proceeds as follows. We first extract a subsequence of unidimensional grid points $S_1, S_2, ...$ from the extrema of the Chebyshev polynomials satisfying $|S_1| = 1$ $|S_i| = 2^{i-1} + 1$ for $i > 1$ and $S_i \subset S_{i+1}$. The first such four nested sets are:

$$
\begin{aligned}
&S_1 = \{0\} \\
&S_2 = \{0, -1, 1\} \\
&S_3 = \left\{0, -1, 1, \frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right\} \\
&S_4 = \left\{0, -1, 1, \frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{-\sqrt{2+\sqrt{2}}}{2}, \frac{-\sqrt{2-\sqrt{2}}}{2}, \frac{\sqrt{2-\sqrt{2}}}{2}, \frac{\sqrt{2+\sqrt{2}}}{2}\right\},
\end{aligned}
$$

equal to the extrema of the 1st, 3rd, 5th, and 7th Chebyshev polynomials of the 1st kind.

To form multidimensional grid points, we can take d-fold products of the unidimensional sets above. In particular, let:

$$
\mathcal{K}^{\boldsymbol{i}} = \prod_{j=1}^{d} S_{i_j}
$$

for $\boldsymbol{i} = (i_1, ..., i_d)$. Finally, let $\mu \geq 1$ be the order of the approximation. Then, the Smolyak sparse grid is formed as:

$$
\mathcal{H}^{d,\mu} = \bigcup_{d \leq |\boldsymbol{i}| \leq d+\mu} \mathcal{K}^{\boldsymbol{i}}
$$

where $|\boldsymbol{i}| = i_1 + \cdots + i_d$. This methodology can easily be extended to construct sparse grids on arbitrary hyperrectangles.

In practice, we employ the Sparse Grids Matlab Kit developed by Piazzola and Tamellini (2022) to construct the Smolyak sparse grids. This toolbox allows one to create a sparse grid of arbitrary order on any given high-dimensions hyperrectangles.

4.3 Monte-Carlo Simulation and Global Solution Algorithm

We use the global solution to determine equilibrium mortgage pricing along the impulse response. Mortgage pricing takes the form of mortgage points:

$$W_t = \mathbb{E}_t \left[e^{-r(\tau - t)} \max\{M_t - \psi R P_\tau, 0\} \right].$$

The difficulty is that the mortgage points depend (nonlinearly) on the equilibrium house price function, but of course equilibrium house prices depend on equilibrium mortgage points.

We therefore follow an iterative procedure, in conjunction with Monte-Carlo simulation, to solve for the global solution. Let $W^0\left(\mathcal{H}^{d,\mu}\right)$ be an initial guess for equilibrium mortgage points on the sparse grid $\mathcal{H}^{d,\mu}$ of model state variables. Then at iteration j, given prices $P^j\left(\mathcal{H}^{d,\mu}\right)$ and investment $I^j\left(\mathcal{H}^{d,\mu}\right)$ on the sparse grid:

1. Use regression to construct low-dimensional polynomial approximations to the house price and investment functions $\hat{P}^j$ and $\hat{I}^j$ with coefficients $\boldsymbol{c}_P^j, \boldsymbol{c}_I^j$.

2. At each point of the sparse grid $\boldsymbol{x} \in \mathcal{H}^{d,\mu}$, use Monte-Carlo simulation with N trials to simulate house prices forward.

   (a) Compute the initial price, investment rate, and mortgage balance based on the assumed LTV ratio.

   (b) Simulate dividends and beliefs forward using Euler-Maruyama method for the SDE system. Specifically, given current dividend $D_0 = D(\boldsymbol{x})$ and lender belief $m_0 = m(\boldsymbol{x})$, the SDE system is:

   $$\begin{aligned} dD_t &= m_t D_t dt + \sigma_D D_t dB_t \\ dm_t &= \vartheta\left(\bar{\mu} - m_t\right) dt + K dB_t. \end{aligned}$$

   Lender beliefs govern the anticipated dividend growth rate. Furthermore, lenders think that their beliefs will follow a standard Kalman filter process, not internalizing that their own beliefs are biased by diagnosticity.

   (c) Use the investment function approximant $\hat{I}^j$ to simulate forward the housing stock as $dH_t = \hat{I}_t^j H_t dt$ and the house price approximant $\hat{P}^j$ to construct house prices at each time step of the simulation.

3. At each point of the sparse grid $\boldsymbol{x} \in \mathcal{H}^{d,\mu}$, compute:

   $$W\left(\boldsymbol{x}\right) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_P \left[e^{-r\tau} \max\{M\left(\boldsymbol{x}\right) - \psi R P_\tau, 0\} \right],$$

   where the expectation $\mathbb{E}_P\left[\cdot\right]$ is conditional on the simulated future house prices for that Monte-Carlo trial. In practice, we compute this by assuming a far out terminal date for the mortgage and solving for the break-even points via recursive backwards induction. This gives $W^{j+1}\left(\mathcal{H}^{d,\mu}\right)$.

4. Given the current solution $W^j\left(\mathcal{H}^{d,\mu}\right)$, solve for the equilibrium price function $P^{j+1}\left(\mathcal{H}^{d,\mu}\right)$ and equilibrium investment function and $I^{j+1}\left(\mathcal{H}^{d,\mu}\right)$ on the sparse grid by equating supply and demand.

5. If $\left\|P^{j+1}\left(\mathcal{H}^{d,\mu}\right) - P^j\left(\mathcal{H}^{d,\mu}\right)\right\| < \varepsilon$ for some specified tolerance level $\varepsilon > 0$, then terminate and use regression to construct the low-dimensional polynomial approximant $\hat{W}^*$ with coefficient $\boldsymbol{c}_W^*$. If not, move to iteration $j+1$ and return to Step 1.

This procedure has the advantage of being highly parallelizable. In practice, it should be noted that the mortgage points have only negligible impact on the quantitative equilibrium impulse response functions. This is emphasized in Appendix C.6, where we allow lenders to have perfect foresight along the impulse response. See in particular Figure C.1. Mortgage points rise dramatically relative to the baseline case since lenders see the crisis coming, but the impact on the equilibrium price path is quite small.

5 Simulated Method of Moments Estimation Procedure For Section 5

We run our simulated method of moments on the Boston University Shared Computing Cluster (SCC) using Matlab in five steps. First, we grid over parameters to find a coarse optimum for quartile four. We use the cluster for this step to break a large grid with millions of parameter combinations into thousands of small chunks that can be run independently and in parallel on separate nodes (we use over 1,000 at one time) and then stitched together. Second, we refine to a finer grid and find a fine optimum for quartile four; this grid is the grid pre-programmed into our replication code. Third, we run fine grids for the other quartiles holding fixed some of the parameters estimated in quartile four; these grids are pre-programmed into our replication code as well. To economize on computing resources, we run the model without priced mortgages for the grids, so before moving on, in a fourth step we check that the model with priced mortgages is very close to the model without priced mortgages. This is the case for all but quartile one, so we run another grid for quartile one (also pre-programmed) with priced mortgage. Fifth, we calculate rents (which is quite computationally intensive) for the optimal parameters for each of the four quartiles and make our final tables and figures.

We include in the replication package not only the Matlab code but also the files used to run the Matlab files. In particular the BU SCC uses a batch scheduler, and we include the .sh "shell" files used to submit jobs to the SCC cluster.

5.1 Step-By-Step Guide

The step-by-step guide to run this on the SCC (or a similar server is) is:

1. Create a folder called "SMM" on your server. Upload the "Template,", "data_moments", and "shared_libraries" folders as well as the .sh files in the "shell_files" folder to the

SMM folder.

- The "Template" folder is the folder that is copied for each chunk and contains all the code to run the simulations for each parameter combination.
- The "shared_libraries" folder contains shared Matlab computational libraries used to solve the model.
- The "data_moments" folder contains the data moments we load in to set some parameters exogenously.
- You will need to update the directory paths in the .sh files.
- It is useful to run build_mex.m in matlab now in the "Template" directory, which will set up the mex files for parallelization on your server (you should also do this in "model_main" on your local computer.

2. Run "smm_nopoints.sh"'. This submits 3,000 jobs to the server, each of which analyzes a chunk of parameter combinations. It also runs the collector when everything is done running. The collector's output is "smm_combined.mat." We have included the parameter grid we used for our fine final run; if you change something you may need to adjust the parameter grid which you can do by finding it in "run_smm_nopoints.m."

3. When the collector is done, download "smm_combined.mat" to the "model_main" folder on your local machine and run "smm_moments.m". (Note that "model_main" includes "smm_combined.mat" after we ran things on the server for our final grid.) This loads the moments from the "data_moments" folder, finds the coarse optimum, and saves it in "optimal_params.mat." At this point, if you change parameters, you may need to adjust the grids in "run_smm_nopoints.m" and re-run until the coarse optimum is in the interior of the grid. To clean things up, delete all the "smm_[number]" folders that were created by "run_smm_nopoints.m." Also rename "smm_combined.mat" on the server.

4. Run "smm_nopoints_xc.sh." This submits 3,000 jobs to the server, each of which analyzes a chunk of parameter combinations for the cross-sectional version (quartiles 1-4), holding ϕ, κ, and ρ fixed at the values estimated for quartile four. It also runs the collector when everything is done running. We have included the parameter grid we used; if you change something you may need to adjust the parameter grid which you can do by finding it in "run_smm_nopoints_xc.m." If you have a different optimum from the one we found in step two, you will need to update ϕ, κ, and ρ in "run_smm_nopoints_xc.m" as these are hard coded.

5. When the collector is done, rename the output "smm_combined.mat" to "smm_combined_xc.mat." Download it to the "model_main" folder on your local machine and run "smm_moments_xc.m." (Note that "model_main" includes "smm_combined_xc.mat" after we ran things on the server for our final grid.) This loads the moments from the "data_moments" folder, finds the coarse optimum, and

saves it in "optimal_params_xc.mat." At this point, if you change parameters, you may need to adjust the grids in "run_smm_nopoints.m" and re-run until the coarse optimum is in the interior of the grid.

6. Run "figures_pointscompare.m" in the "model_main" folder. This creates a plot that compares the model without priced mortgages with the model with priced mortgages to see if we need to run a grid with priced mortgages. We find that the unpriced mortgage model deviates very little from the priced mortgage model except in quartile 1.

7. Run "smm_points_xc.sh." This submits 3,000 jobs to the server, each of which analyzes a chunk of parameter combinations for quartile 1 only with priced mortgages. We do not parallelize the calculation of points, so this runs very slowly.

8. When the collector is done, rename the output "smm_combined.mat" to "smm_combined_xc_q1.mat." Download it to the "model_main" folder on your local machine and run "smm_moments_xc_q1.m." (Note that "model_main" includes "smm_combined_xc_q1.mat" after we ran things on the server for our final grid.) This loads the moments from the "data_moments" folder, finds the coarse optimum, and overwrites "optimal_params_xc.mat" with the priced mortgage optimum for q1. At this point, if you change parameters, you may need to adjust the grids in "run_smm_points.m" and re-run until the coarse optimum is in the interior of the grid.

9. Copy the "Template" folder on the server and name it "postgrid_rents." Upload "optimal_params_xc.mat" to the folder and then Run "postgrid_rents.sh" on the server. This calculates the rents, which is computationally intensive, on the server and creates a file called "postgrid_rents.mat" that includes the calculated rents. When it is finished, download this to the "model_main" folder on your local machine.

10. Create the tables and figures by running: "figure_main.m" which creates the tables and figures in the text and "figures_pf.m" which creates the lender perfect foresight figure in the appendix, all of which are saved in "model_main/output." In particular:

    - moments.text, which is the bottom half of Table 2. The top half is entered manually based on "optimal_params_xc.mat"
    - paper_crossicites.pdf, which is Figure 7.
    - paper_rents.pdf, which is Figure 8.
    - paper_structuralcombined.pdf, which is Figure 9.
    - slides_lenderforesight.pdf, which is Figure C.1.

5.2 Ancillary Calibration Targets

Most of the calibration targets are loaded from csv files created in Stata as described in section 1 (these can be found in the "data_moments" folder or hard coded.

Two sets of calibration targets are worth further discussion. First, the repeat sales residual standard deviation from DataQuick is used as a moment in our calibration procedure. As described in Section 5.3 of the paper, we load all of the DataQuick deeds data from 1988-2013 for non-distressed ales of single family homes and condominiums and run a repeat sales regression of log price on house fixed effects and census tract-by-quarter fixed effects and use the residuals to create this moment. The do file to run this analysis is "dq_residuals.do" in the "model/data_moments" folder. We unfortunately cannot share the underlying data, which was obtained from DataQuick by Harvard University. This data set and its construction are detailed in Online Appendix B of Guren, 2018. Second, we load the initial loan balance distribution from the 1995 Survey of Consumer Finances. This file "ltv-scf-1995.csv" in the "data_moments" folder is created by the do files described in Section 1. The "import_LTV.m" file in "model_main" loads this and saves it in "LTV_DATA", which is used by the rest of the model files.

Finally, we use data from Shiller and Thompson (2022) for our expectations moments. We use the average expected annual house price growth over the next 10 years for the three counties from quartile 4 in their data set (Alameda, CA; Middlesex, MA; and Orange, CA). These are 7.7, 7.2, and 8.1 in 2006 yielding an average of 7.7% and 6.2, 4.2, and 4.8 in 2019 yielding an average of 5.1%. We then adjust for inflation by subtracting average expected inflation over the next 5 years from the Michigan Survey, which is 2.7% in 2006 and 2.0% in 2019. This yields 5.0% real expected price growth in 2006 and 3.1% real expected price growth in 2019, which are the two moments we feed into the SMM.

6 Model Replication Package File By File Summary

In this section, we provide a a description of each of the Matlab files used in the simulation of our model. These files are found in the folder "model/model_main."

- `BGP_Rates.m`: Constructs house price appreciation and construction rates on initial and final BGP
- `build_mex.m`: Constructs the mex file of `MC_loop_parallel.m` for speed optimization
- `calc_moments.m`: Takes the price and foreclosure paths and calculates the SMM moments.
- `calc_points_pf.m`: Calculates mortgage points in the case of perfect lender foresight along the impulse response
- `cheb_1st.m`: Evaluates Chebyshev polynomial of the 1st kind
- `cheb_1st.m`: Evaluates Chebyshev polynomial of the 2nd kind
- `cheb_create.m`: Returns Chebyshev interpolant coefficients of loan distribution

- `collect_smm.m`: This is the "collector" file that loops over the results of the different SMM blocks run separately on different nodes on the cluster and then combines them into one file for analysis.
- `computeRentsNew.m`: Returns equilibrium rents along the impulse response path
- `ConstructApprox.m`: Constructs polynomial approximation of mortgage points function
- `create_sparsegrid.m`: Constructs a Smolyak sparse grid of given order
- `figures_main.m`: This file creates the final figures for the paper after the SMM procedure is complete.
- `figures_pf.m`: This file creates the lender perfect foresight figure in the appendix.
- `gen_BGP_LTV_dist.m`: Constructs steady-state LTV distribution
- `gen_mex_MC_parallel.m`: Helper function used in `build_mex.m`
- `gen_parameters_passparams.m`: Initializes parameters, constructs the initial LTV distribution, and solves for the initial price
- `genZ.m`: Produces matrix of Gaussian shocks to be used for Monte-Carlo simulation
- `grid_construct.m`: Constructs a sparse grid for global solution
- `import_LTV.m`: Imports the initial LTV distribution from the SCF data
- `LTV_DATA.mat`: The imported LTV distribution from the SCF
- `MC_loop_parallel.m`: Monte-Carlo evaluation of mortgage points pricing
- `MC_loop_parallel.mex`: Mex file of `MC_loop_parallel.m`
- `optimal_params.mat`: The optimal parameters for quartile 4 from the SMM procedure
- `optimal_params_xc.mat`: The optimal parameters across the 4 quartiles from the SMM procedure
- `PI_System_BGP.m`: Supply and demand system for initial BGP
- `PI_System_Global.m`: Supply and demand system for global solution
- `PI_System_Foreclosure.m`: Supply and demand system for impulse response
- `postgrid_rents.m`: Calculates rents using the final SMM optimal parameters.
- `PriceBeliefsCalc.m`: Solves for house price forecasts on equilibrium impulse response at specified times

- `PriceBeliefsCalcFull.m`: Solves for house price forecasts along entire equilibrium impulse response
- `PV.m`: Constructs present value of dividends
- `run_model_smm_nopoints_beliefs.m` Main function which returns equilibrium impulse response as well as price forecasts on the impulse response but sets mortgage points to zero
- `run_model_smm_beliefs.m` Main function which returns equilibrium impulse response as well as price forecasts on the impulse response
- `run_model_smm_beliefs_nomc.m` Single threaded version of run_model_smm_beliefs.m (for use on a single-core node)
- `run_model_smm_pf.m` Main function which returns equilibrium impulse response when lenders have perfect foresight
- `run_model_smm_rents.m` Main function which computes equilibrium rents along the impulse response path
- `run_smm.m`: Runs the Q4 SMM with points
- `run_smm_nopoints.m`: Runs the Q4 SMM without points
- `run_smm_xc.m`: Runs the cross-sectional (all 4 quartiles) SMM with points
- `run_smm_nopoints_xc.m`: Runs the cross-sectional (all 4 quartiles) SMM without points
- `simulate_impulse_forecast.m` Simulates equilibrium impulse response function
- `simulate_impulse_pf_new.m` Simulates equilibrium impulse response function when lenders have perfect foresight
- `smm_moments.m`: Loads the target moments, loads the combined runs from the server in smm_combined.mat, and runs the quartile 4 SMM optimization, saving the output in optimal_params.mat
- `smm_moments_xc.m`: Loads the target moments, loads the combined runs from the server in smm_combined.mat, and runs the cross-sectional SMM optimization, saving the output in optimal_params_xc.mat
- `smm_moments_xc_q1.m`: For quartile 1 only we have that the nopoints version is not close to the points version. So redoes the smm_moments_xc analysis for Q1 only on the points version

- `Solve_Eq_Price_BGP.m`: Solves for the equilibrium house price and construction rate on the initial BGP
- `Solve_Eq_Price_Global.m`: Solves for the equilibrium house price and construction rate for the global solution
- `Solve_Eq_Price_Impulse.m`: Solves for the equilibrium house price and construction rate on the impulse response
- `solve_global.m`: Global solution for equilibrium points
- `solve_global_nomc.m`: Non-multi-threaded version of solve_global.m

The replication package also contains some additional software packages which the main code makes use of. These are in "model/shared_libraries". The libraries are "chebfun," which provides code to compute Chebyshev interpolants, and "sparsegrids," which provides code to compute Smolyak sparse grids. These are developed respectively by Driscoll et al. (2014) and Piazzola and Tamellini (2022). We also use subtightplot.m for plots, which we include in the replication package.

Finally, the shell files used to run the server and described above are in model/shell_files.

References

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra (2019). "Regional Heterogeneity and the Refinancing Channel of Monetary Policy". *The Quarterly Journal of Economics* 134 (1): 109–183.

Consumer Financial Protection Bureau (2022). "Home Mortgage Disclosure Act Data". Online at https://www.consumerfinance.gov/data-research/hmda/historic-data/, accessed June 2020.

CoreLogic (2018). "CoreLogic CBSA-Level House Price Index, Tier 11". Proprietary Data Obtained From Vendor.

Crone, Theodore M, Leonard I Nakamura, and Richard Voith (2010). "Rents Have Been Rising, Not Falling, in the Postwar Period". *The Review of Economics and Statistics* 92 (3): 628–642.

Driscoll, Tobin A, Nicholas Hale, and Lloyd N Trefethen (2014). *Chebfun guide*.

Eckert, Fabian, Teresa Fort, Peter Schott, and Natalie Yang (2020). "Imputing Missing Values in the US Census Bureau's County Business Patterns". NBER Working Paper No. 26632.

Federal Housing Finance Agency (2020). "Five-Digit ZIP Code Annual House Price Indices (Developmental Index; Not Seasonally Adjusted)". Online at https://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_AT_BDL_ZIP5.xlsx, accessed November 2020.

Federal Reserve Bank of St. Louis (2022). "FRED Economic Data". Online at https://fred.stlouisfed.org/, accessed October 2022.

Federal Reserve Board of Governors (1995). "Survey of Consumer Finances for 1995". Online at https://www.federalreserve.gov/econres/files/scfp1995s.zip, accessed April 2021.

Freddie Mac (2021). "Freddie Mac House Prince Index". Online at https://www.freddiemac.com/research/indices/house-price-index, accessed February 2021.

Gao, Zhenyu, Michael Sockin, and Wei Xiong (2020). "Economic Consequences of Housing Speculation". *The Review of Financial Studies* 33 (11): 5248–5287.

Guren, Adam M. (2018). "House Price Momentum and Strategic Complementarity". *Journal of Political Economy* 126 (3): 1172–1218.

Gyourko, Joseph, Albert Saiz, and Anita Summers (2008). "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index". *Urban Studies* 45 (3).

Judd, Kenneth L, Lilia Maliar, Serguei Maliar, and Rafael Valero (2014). "Smolyak method for solving dynamic economic models: Lagrange interpolation, anisotropic grid and adaptive domain". *Journal of Economic Dynamics and Control* 44: 92–123.

Larson, William, Jessica Shui, Morris Davis, and Stephen Oliner (2021). "The Price of Residential Land for Counties, ZIP Codes, and Census Tracts in the United States". *Journal of Monetary Economics* 118 (March): 413–431.

Logan, John, Zengwang Xu, Brian Stults, and Charles Zhang (2022). "The Longitudinal Tract Data Base". Online at https://s4.ad.brown.edu/Projects/Diversity/Researcher/Bridging.htm, accessed June 2022.

Lutz, Chandler and Ben Sand (2019). "Highly Disaggregated Topological Land Unavailability". Working Paper, York University.

Piazzola, C. and L. Tamellini (2022). "The Sparse Grids Matlab kit - a Matlab implementation of sparse grids for high-dimensional function approximation and uncertainty quantification". *ArXiv* (2203.09314).

Ruggles, Steven, Sarah Flood, Matthew Sobek, Brockman Danika, Grace Cooper, Stephanie Richards, and Megan Schouweiler (2022). "IPUMS USA". Online at usa.ipums.org, accessed various dates.

Saiz, Albert (2010). "The Geographic Determinants of Housing Supply". *The Quarterly Journal of Economics* 125 (3): 1253–1296.

Shiller, Robert J. and Anne K. Thompson (2022). "What Have They Been Thinking? Homebuyer Behavior in Hot and Cold Markets: A Ten-year Retrospect". *Brookings Papers on Economic Activity* 2022: 307–344.

Statisticians of American Religious Bodies, Association of (1992). "Churches and Church Membership in the United States, 1990". Distributed by the Association of Religion Data Archives online at https://www.thearda.com/data-archive?fid=CMS90CNT, accessed February 2021.

United State Census Bureau (1993). "1993 State Local Government Finance Historical Datasets and Tables". Online at https://www.census.gov/data/datasets/1993/econ/local/public-use-datasets.html, accessed February 2021.

— (2022). "County Population Estimates". Online at https://www.census.gov/programs-surveys/popest.html, accessed October 2022.

United States Census Bureau (2011). "USA Counties: 2011, Land Area". Online at https://www.census.gov/library/publications/2011/compendia/usa-counties-2011.html#LND, accessed November 2020.

— (2020). "County Business Patterns". Online at https://www.census.gov/programs-surveys/cbp/data.html, accessed August 2020.

— (2022). "Annual Estimates of Housing Units for Counties in the United States". Online. 2000-2010 data at https://www2.census.gov/programs-surveys/popest/tables/2000-2010/intercensal/housing and 2010-2019 data at https://www2.census.gov/programs-surveys/popest/tables/2010-2019/housing/totals/CO-EST2019-ANNHU.xlsx, accessed May 2022.

United States Department of Agriculture (2022). "Natural Amenities Scale". Online at https://www.ers.usda.gov/data-products/natural-amenities-scale/, accessed May 2022.

University of Michigan Surveys of Consumers (2022). "Table 33: Expected Change in Prices During the Next 5 Years". Online at https://data.sca.isr.umich.edu/data-archive/mine.php, accessed July 2022.

Zillow (2022). "Zillow Home Value Index". Online at https://www.zillow.com/research/data/, accessed July 2021 and June 2022.