Anatomy of the Beginning of the Housing Boom Across U.S. Metropolitan Areas

Fernando Ferreira and Joseph Gyourko
The Wharton School, University of Pennsylvania
& NBER

Online Appendix

November 11, 2021

This online appendix provides added detail on the data used in our paper "Anatomy of the Beginning of the Housing Boom Across U.S. Metropolitan Areas", as well as various robustness analyses related to that work.

I. Data Description

The home purchase and financing transactions files compiled by the data vendor DataQuick (now CoreLogic) are the foundation of the rich micro data used in this paper. They permit us to observe sales transactions of single family units and homes in condominium or multi-unit structures. We also observe the financing associated with those purchases, as well as subsequent refinancings and subordinate mortgages. Our sample includes this information for the 94 metropolitan areas listed in Online Appendix Table 3 below. Different metropolitan areas enter the sample at different times, some as early at 1993(Q1), so homes purchased before these dates do not enter our study sample (unless they are resold later). The majority of MSAs have data available since 1995, and all MSAs have data since 1998.

Detailed information is provided on the following variables (among others): (a) transaction date; (b) name of the buyer if the observation is for a purchase; name of the owner if the observation is for a refinancing or other debt; (c) name of the seller if the observation is for a home purchase; (d) names of up to three lenders for any type of transaction involving new debt; (e) sales price for all home purchases; (f) mortgage amounts for up to three loans on all observations using any type of financing; (g) street address and census tract of the underlying home; (f) various home characteristics including age of the home, size as reflected in the number of bedrooms, bathrooms, and square footage, etc. and (g) codings provided by CoreLogic indicating whether a transaction involves a home being foreclosed by a creditor, as well as whether the home is being sold out of foreclosure to a new owner; in both cases, names of the principals are reported, along with a purchase price for the latter type of transaction.

Because individual owners and all their financings can be tracked over time, we use these data to create a panel of individual ownership sequences. An ownership sequence is the complete span of time a unique owner owns a given residence. Our final panel contains 33,545,252 ownership sequences on 19,648,475 homes.

A. The Number and Types of Transactions

The predominant type of transaction is an arms-length purchase of an existing home. These constitute 80.2% of all our home sales transactions. Arms-length sales of new homes from the builder (or other entity) to a household make up another 11.2% of all purchases. The remaining sales observations are comprised of purchases out of foreclosure (8.6%). CoreLogic does not code these as arms-length trades between two disinterested parties, but they are readily identifiable from another variable categorizing 'distress' transactions.[1]

We also observe about 48 million financings not associated with a home purchase. These include refinancings and the taking on of junior debt. First, second and third loans at purchase are clearly identified. However, CoreLogic does not identify whether a subsequent financing within a unique ownership sequence represents a refinancing of existing debt or the taking on of an additional loan. We adopt the following rule to distinguish between the two cases. If a new mortgage taken out subsequent to purchase has an initial loan balance that is more than 50% of the total mortgage balance taken out at purchase or is more than 50% of the imputed current price of the home, we assume the new loan is a refinancing that replaces the prior debt; otherwise, it represents junior debt, which is added to the outstanding loan balance. Using this rule, we observe about 34 million refinancings and just over 14 million second loans.

B. Classifying Owners

Each ownership sequence is classified as one of five types based on the type of financing used by the owner. The most straightforward is those who buy their housing unit without using any debt. These are referred to as Cash owners in all tables and figures. They constitute a relatively stable 10%-11% share of our sample until 2010, after which their share increases to over 16% in 2012. If an owner purchases a house with no debt, but subsequently takes out a mortgage, that owner is no longer considered a Cash owner as of the quarter of the loan origination.

All other ownership sequences involve the use of some type of debt. We divide each of these owners into one of four groups of borrowers: (a) Prime; (b) Subprime; (c) FHA/VA-

[1] The seller in these cases typically is some type of financial entity, while the buyer usually is a household. See the discussion below for more on these transactions. Some do not consider these 'normal' sales, but they certainly are home purchases, and we count them as such. Their transaction prices also are included in the price series described below, although we can do all our analysis excluding them.

insured; or (d) 'Small'. Lender lists are used to define subprime mortgages because we do not have access to credit score micro data. More specifically, we define a borrower as subprime if it obtained its loan(s) from a lender on either the HUD or *Inside Mortgage Finance* lists, but the loan was not insured by FHA or VA. This group is called Subprime in all tables and figures.[2] The subprime group has very high rates of home loss, which is consistent with the rest of the literature regardless of their data and procedures for distinguishing subprime from prime.

However, we do not categorize all other borrowers as Prime. Two other categories are included to help ensure we do not conflate subprime and prime owners. The first is comprised of borrowers whose loans were guaranteed by FHA or VA (regardless of lender identity). They are labeled FHA/VA owners in all tables and figures.[3] These loans often are considered of subprime quality because of the very high initial loan-to-value ratios usually involved, but we treat them separately from the 'private' subprime group.

All remaining owners with debt are Prime borrowers by definition. Their share always exceeds 50%, and it rose, not fell, as the boom built, from a low of 54.9% in 2000(Q1) to a high of 65.6% in 2008(Q1). Thus, the rough doubling of Subprime share over the same period is at the expense of the FHA/VA-insured sector, not the Prime sector.

C. Constant Quality House Prices

We begin by creating a MSA-level (m) constant quality house price series by quarter (t) using hedonic regressions. Price (*HP*), in logarithmic form, is modeled as a function of the square footage (*Sqft*) of the home entered in quadratic form, the number of bedrooms (*Bed*), the number of bathrooms (*Bath*), and the age of the home (*Age*).[4] The hedonic index values are

[2] The entities on the subprime lender list generally distinguish among the several units of a lender. For the HUD list in particular, identification was based on the HMDA identification number of the entity, and different subsidiaries of a large bank typically had different ID numbers. Thus, having a subsidiary of (say) Bank of America that HUD believes specializes in subprime lending on the list does not mean that all of Bank of America's mortgage issuance gets classified as subprime. Banks and subsidiaries also enter and leave the HUD list over time. The HUD list also ends in 2005. The *Inside Mortgage Finance* publication also lists specific units of some large financial institutions, but we also consider those units as subprime if they ever show up on that publication's list.

[3] Ten metropolitan areas in the northeastern part of the country do not report data for this particular variable. They are Barnstable Town, MA, Boston-Cambridge-Quincy, MA-NH, Bridgeport-Stamford-Norwalk, CT, Harford-West Hartford-East Hartford, CT, New Haven-Milford, CT, Pittsfield, MA, Providence-New Bedford-Fall River, RI-MA, Springfield, MA, and Worcester, MA. We still include observations from these metropolitan areas in our regression analysis, but code this variable for them so that it is estimated separately from that for the other MSAs.

[4] The hedonic regression in Equation (1) contains a number of categorical variables created to control for differences in housing quality. Separate vectors were created for the number of bedrooms (Bed), the number of bathrooms (Bath) and the age of the home (Age). In the case of bedrooms, ten dichotomous dummies were used to control for

derived from the coefficients in the vector α_6 on the year-quarter dummies (*YearQtr*) in the following equation:

$$(1)\ Log(HP_{m,t}) = \alpha_0 + \alpha_1 * Bed_{m,t} + \alpha_2 * Bath_{m,t} + \alpha_3 * Age_{m,t} + \alpha_4 * Sqft_{m,t} + \alpha_5 * Sqft^2_{m,t} + \\ + \alpha_6 YearQtr_t + \epsilon_{m,t},$$

where $\epsilon_{m,t}$ is an idiosyncratic error term. The estimated indexes are then normalized to 100 in 2000(Q1) for all MSAs.

Our use of hedonic price indexes is in contrast to the now widespread use of repeat sales indexes, which were reintroduced and popularized by Case and Shiller (1987). We employ hedonic price indexes because their data requirements are much less onerous. To check the robustness of our estimates, we compared the hedonic index with the S&P/Case-Shiller repeat sales index, and a repeat sales index we created using the CoreLogic files for the Las Vegas-Paradise metropolitan area.[5] While there are some small differences in the early- to mid-1990s, all three indexes capture the remarkable boom and bust in this market – with the hedonic method only slightly understating the peak of the boom. We do observe small differences in the short-run volatility of the price level indexes, presumably because Case-Shiller adopts additional procedures to smooth its index. Even so, the simple correlation across any two price indexes is 0.99. Correlations among price appreciation rate series are almost as high (0.97). Similar results are found for most other markets that are comparable with Case-Shiller.[6] Hence, our hedonic

---

the number of bedrooms ranging from less than 1 (which includes 0 and 0.5 bedrooms in the raw data) to a top code of 9 for homes with nine or more bedrooms. In this case, each dummy represented a unit increase in the number of bedrooms (e.g. there are dichotomous dummies created for homes with <1, 1, 2, 3, 4, 5, 6, 7, 8, and 9+ bedrooms). In the case of bathrooms, we included controls for homes with fewer than 1 bathroom (again, 0 or more typically, 0.5 bathrooms), a top code for units with seven or more bathrooms, dummies for each half unit increase from 1 through 5, and then controls for each unit increase until seven. More specifically, the twelve categories were: <1, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, and 7+. There are nine categories of the Age vector from Equation (1). They range from newly built homes with an age of zero to homes at least 40 years old. The specific age categories are as follows: 0, 1, 2-5, 6-9, 10-14, 15-19, 20-29, 30-39, and 40+ years old. The other quality control in the hedonic estimation, the square footage of the living space in the home, is continuous in nature and was entered in quadratic form as noted in Equation (1).

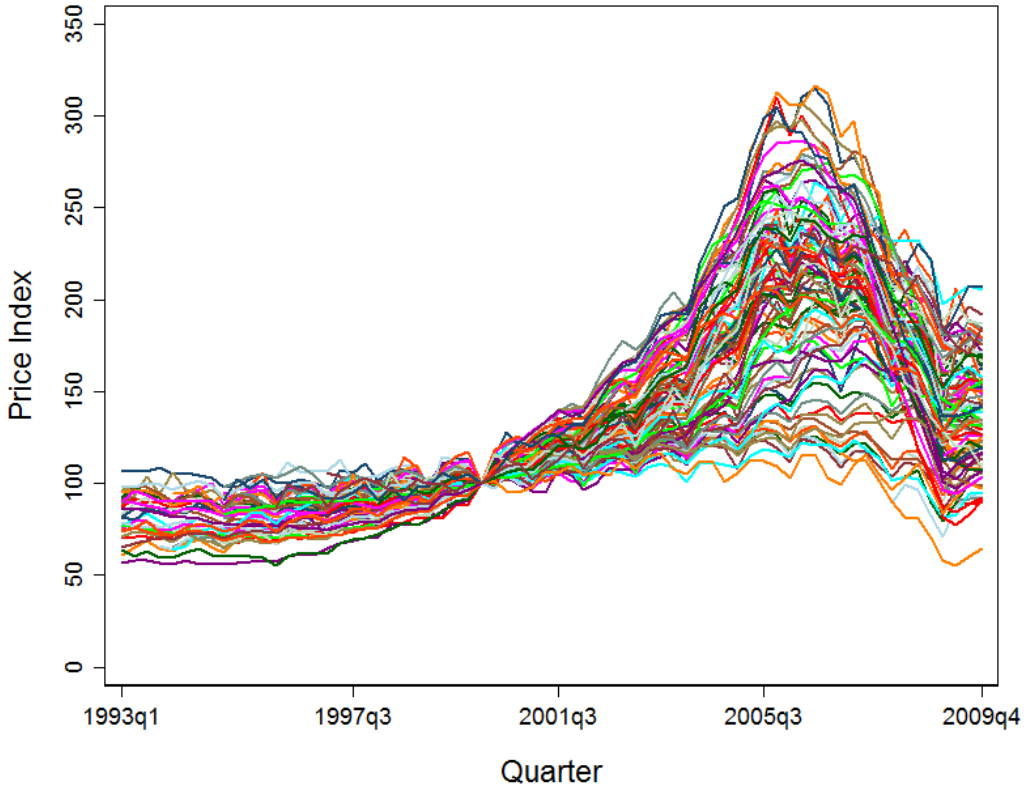
[5] Our repeat sales index uses a subsample of houses that had at least two transactions recorded in DataQuick. The index created does not include observations on units that transacted multiple times within a 10-day period. We also account for the most recent major renovation in a house, as that sequence of transactions is unlikely to have similar housing features over time.

[6] Of the 14 Case-Shiller markets that overlap with the CoreLogic files, the least good match between our hedonic index with the S&P/Case-Shiller repeat sales index is for Cleveland. The correlation between our hedonic index and the Case-Shiller repeat sales index is only 0.61 for that market. The simple correlation of appreciation rates on the two different indexes based on DataQuick is higher at 0.87, but still below that for all other comparable markets.

method captures the price movements tracked with other widely used methods to a very great extent.

The constant quality price indexes used throughout our analysis are plotted below in Online Appendix Figure 1:

Online Appendix Figure 1: Individual Metropolitan Area hedonic price indexes by quarter



Notes: Each line represents a hedonic price index that was separately estimated for each MSA. The index for 2000Q1 is normalized to 100 for each MSA.

E. Leverage at Purchase and Over Time

Loan and purchase price data are combined to compute loan-to-value (LTV) ratios. Doing so at purchase is straightforward: divide the sum of all mortgages taken out at purchase by the purchase price recorded by CoreLogic. FHA/VA-insured loans have much higher initial LTVs (close to 1) than both prime and subprime loans throughout our full sample period, and actually fell slightly over our sample period. Subprime borrower average initial LTVs did increase from about 81% to 85% as the boom built in the mid-2000s. There is a more modest

increase in Prime borrower initial LTVs over the same time period. Thus, there was not a dramatic surge in initial leverage ratios for the typical borrower in any sector of the mortgage market while the long boom in house prices built.

Current LTV by quarter must be estimated. Fortunately, in addition to having panels of ownership sequences that make its estimation feasible, two features of our data allow for a more accurate estimation than exists in other research: (a) the complete history of home financings, including refinancings and second loans; and (b) neighborhood-level house price indexes.[7] In imputing the numerator, we presume that all new debt taken on is fully amortizing, 30-year, fixed rate product. This is a conservative assumption that almost certainly leads us to understate true LTV, particularly on subprime product which the literature suggests more often involved adjustable rate mortgages (ARMs) and terms that did not require immediate amortization of principal. To impute current house value in the denominator, we start with house price at purchase, and update it on a half-year basis using our neighborhood-level price indexes. Noise in the denominator can arise in different ways. For example, values for distressed properties are likely to be overstated because they probably were receiving lesser maintenance and repair-related investment. This provides another reason why current LTV could be underestimated. However, we suspect that variation provided by refinancings, second loans and the local price index likely overshadow the measurement error due to this factor.

F. Identifying Speculators

Researchers and popular commentators have argued that speculators may have played an important role in the building of the last housing boom, thereby helping make its ultimate demise worse (e.g., Haughwout, et. al. (2011); Chinco and Mayer (2014)). We identify speculators in one of two ways. First, we follow Chinco and Mayer (2014) who reasoned that since speculators would not be living in the purchased unit, they would have their tax bills sent to another address. We compare the precise street address of the housing unit with the address to which the tax bill is sent – the 'Tax Address' in the DataQuick files. Whenever the two are appreciably different, we

[7] Some private data vendors have begun creating cumulative LTVs on observations in their loan-level data sets. Essentially, they do it as we do, by linking to deeds records (which is what CoreLogic does) so they can track a given observation over time. To our knowledge, this has not yet shown up in current or published research.

call that purchaser a speculator.[8] The second way we identify whether a purchaser is a speculator is by whether the buyer has a name that is a business. This includes corporate or commercial names that include LLC or INC in them, homebuilders, or trusts (especially mortgage-backed securities trusts which are typically identified by a four-digit number in their names).[9]

G. Demographics and Income of Borrowers

A weakness of the CoreLogic files is that they do not contain any information on the owners beyond their names. To gain more insight into borrower demographic characteristics (race and gender of the head of the household) and the self-reported income levels, we match individual sales transactions to loan application data in the Home Mortgage Disclosure Act (HMDA) files. Observations are merged as follows. In the first step, each transaction was matched to a loan using the year in which the transaction occurred, the full 11 digit Census tract number, the lender name, and the exact loan amount. In cases where there were multiple matches, one of them was randomly assigned as being a true match while the rest were considered unmatched. The remaining unmatched observations were then merged based only on year, Census tract and exact loan amount with multiple matches being randomly assigned as in the first step. This two-step process was repeated several times allowing for the loan amounts to differ from each other in increments of $1,000 up to a total allowable difference of $10,000. Any observations remaining after this process then went through an identical matching procedure using 9 digit Census tract numbers. Observations surviving that procedure are considered to be unmatched. In total, 92.7% of the sales transactions in CoreLogic were matched at some point in the procedure. Of those, approximately 60% were matched in the first step. Because we are unsure about the quality of the matches in subsequent steps, in the empirical work below we always distinguish the demographics in two groups – perfect and imperfect matches – and include both in the estimation. Reported regression coefficients are for the perfect matches only.

[8] By appreciably different, we generally mean that more than one number in the street address before the zip code differs.

[9] Other academic research has identified speculators by whether they 'flip' properties quickly (e.g., Bayer, Geissler, Magnum and Roberts (2011)). We also investigated those cases, but more than 99% of them were already encompassed by our measures of tax address and names of business.

Finally, the demographic data for Cash buyers is missing by definition because they never took out a loan, and hence, cannot be matched with any HMDA observation.

References

Bayer P., C. Geissler, K. Mangum, and J. Roberts. 2020. "Speculators and Middlemen: The Strategy and Performance of Investors in the Housing Market." forthcoming, *The Review of Financial Studies*

Chinco A. and C. Mayer. 2016. "Misinformed Speculators and Mispricing in the Housing Market." *The Review of Financial Studies* 29(2):486-522

Haughwout A., D. Lee, J. Tracy, and W. van der Klaauw. 2011. "Real Estate Investors, the Leverage Cycle, and the Housing Market Crisis." Federal Reserve Bank of New York Staff Reports.

II. Summary Statistics—Online Appendix Tables 1-3

Online Appendix Table 1 compares the final sample used in our paper to that for the entire country. This is followed in Online Appendix Table 2 with detailed summary statistics on all variables used in the empirical estimations in our paper. Finally, Online Appendix Table 3 provides timelines of all booms in each of the 94 markets studied in our analysis.

Online Appendix Table 1: Representativeness of final sample

	All U.S.	Final
	(1)	(2)
Number of MSAs	355	94
Population of MSAs	676,287	1,391,790
	(1,536,134)	(2,596,025)
% Northeast	0.11	0.11
	(0.31)	(0.31)
% Midwest	0.20	0.08
	(0.40)	(0.27)
% South	0.36	0.24
	(0.48)	(0.43)
% West	0.20	0.49
	(0.40)	(0.50)
% White	0.80	0.78
	(0.12)	(0.11)
% Black	0.10	0.08
	(0.11)	(0.08)
% College Degrees or higher	11.17	12.10
	(3.36)	(3.26)
Median Family Income	$53,882	$58,555
	(9,350)	(10,599)
Median House Value	$164,787	$271,807
	(108,058)	(150,469)

Notes: All sociodemographic data based on Census 2005 (American FactFinder). First column presents averages and standard deviations (in parenthesis) for all MSAs with available data in the country. Column 2 presents descriptive statistics for our final sample of 94 MSAs.

Online Appendix Table 2: Summary Statistics on Variables Used in the Empirical Analysis

Variable Name	N	Mean	SD	Min	Median	Max
Price						
Price Index	6222	134	48	55	121	317
Price Growth Rate	5846	0.029	0.125	-0.508	0.037	0.439
Fundamentals						
HMDA income (1K dollars)	6222	89	32	26	82	249
HMDA income growth rate	5846	0.038	0.135	-0.587	0.042	1.139
White share, HMDA	6222	0.790	0.131	0.367	0.826	1
Minority share, HMDA	6222	0.149	0.106	0	0.125	0.609
BEA Income (1K dollars)	1710	32	10	16	30	106
BEA Income growth rate	1620	0.037	0.035	-0.294	0.041	0.358
Unemployment rate	6222	0.069	0.036	0.016	0.057	0.383
White share, schools	1724	0.609	0.191	0.145	0.646	0.940
Minority share, schools	1724	0.312	0.173	0.035	0.292	0.814
Transaction Characteristics						
Prime share	6222	0.495	0.147	0.098	0.497	0.854
Subprime share	6222	0.098	0.082	0	0.085	0.433
FHA share	6222	0.148	0.126	0	0.129	0.660
Cash-only share	6222	0.214	0.114	0.019	0.187	0.758
Average initial LTV	6222	0.656	0.112	0.176	0.681	0.878
Variable-rate share	6222	0.167	0.169	0	0.111	0.854
Number of refinancing loans	6222	2015	4626	0	499	82814
Number of resale transactions	6222	2580	3906	24	1028	31767

Notes: Each variable is described in the data section.

Online Appendix Table 3. Estimated breakpoints

	One-Break	Two-Break		Three-Break		
MSA Name	BP1	BP1	BP2	BP1	BP2	BP3
(1) Akron, OH	2003q4[§]					
(2) Atlantic City, NJ	2003q1	2001q3	2003q1			
(3) Bakersfield, CA	2002q4	2001q1	2004q2			
(4) Baltimore-Towson, MD	2002q1					
(5) Barnstable Town, MA	1998q4	1999q2	2003q4[§]	1998q1	1999q4	2003q4[§]
(6) Bellingham, WA	2003q3	1997q3	2003q3			
(7) Boston-Cambridge-Quincy, MA-NH	1997q4	1997q4	2003q2[§]	1995q4	1999q3	2003q2[§]
(8) Bremerton-Silverdale, WA	2003q1	1998q1	2003q1			
(9) Bridgeport-Stamford-Norwalk, CT	1997q2	1998q3	2005q4[§]	1997q2	1999q3	2005q4[§]
(10) Cape Coral-Fort Myers, FL	2004q4	2001q2	2004q4			
(11) Carson City, NV	2003q3	2001q1	2004q2			
(12) Chicago-Naperville-Joliet, IL-IN-WI	2006q1[†§]					
(13) Chico, CA	2001q3	2000q3	2003q2			
(14) Cincinnati-Middletown, OH-KY-IN	2003q4[§]	2004q1[§]	2005q1			
(15) Cleveland-Elyria-Mentor, OH	2002q1[§]					
(16) Colorado Springs, CO	2002q2[§]	2002q2[§]	2005q1	2000q4	2002q1[§]	2005q1
(17) Columbus, OH	2003q4[§]					
(18) Corvallis, OR	2004q3					
(19) Dayton, OH	2002q2[†]					
(20) Deltona-Daytona Beach-Ormond Beach, FL	2004q1					
(21) Denver-Aurora, CO	2001q4[§]	2001q3[§]	2002q4[§]			
(22) Detroit-Warren-Livonia, MI	2002q2[§]	2002q2[§]	2003q2			
(23) Eugene-Springfield, OR	2004q2					
(24) Flagstaff, AZ	2004q3					
(25) Fort Collins-Loveland, CO	2002q1[§]					
(26) Fort Walton Beach-Crestview-Destin, FL	2003q4	2002q2	2004q1			
(27) Fresno, CA	2002q4	2001q1	2002q4			
(28) Gainesville, FL	2002q1					
(29) Grand Junction, CO	2004q1					
(30) Hanford-Corcoran, CA	2003q4	2001q1	2003q4			
(31) Hartford-West Hartford-East Hartford, CT	1998q2	1998q1	2000q2	1996q2	1998q2	2000q2
(32) Jacksonville, FL	2002q4	2001q1	2004q2			
(33) Kingston, NY	1998q4[†]					
(34) Lakeland-Winter Haven, FL	2004q3					
(35) Las Vegas-Paradise, NV	2004q1	2004q2	2005q2[§]			
(36) Los Angeles-Long Beach-Santa Ana, CA	1997q4	1997q3	2002q2	1997q3	2002q3	2005q4[§]
(37) Madera, CA	2002q3					
(38) Medford, OR	2003q1	1999q3	2004q1	1999q3	2002q4	2004q1
(39) Memphis, TN-MS-AR	2005q1					
(40) Merced, CA	2000q2	2000q1	2004q2	1998q1	2000q2	2004q2
(41) Miami-Fort Lauderdale-Miami Beach, FL	2001q3					
(42) Modesto, CA	2000q3	1998q2	2000q4	1998q2	2000q4	2004q2
(43) Mount Vernon-Anacortes, WA	2005q3					
(44) Napa, CA	1998q1	1997q3	2000q1	1997q3	2000q1	2001q3[§]
(45) Naples-Marco Island, FL	2004q1[†]					
(46) New Haven-Milford, CT	1998q3	1997q1	2001q1	1996q3	1998q3	2001q1
(47) New York-Northern New Jersey-Long Island, NY-NJ-PA	2002q2	2002q2	2006q1[§]			
(48) Norwich-New London, CT	2001q1	1997q4	2002q2			

	One-Break	Two-Break		Three-Break		
MSA Name	BP1	BP1	BP2	BP1	BP2	BP3
(49) Ocala, FL	2005q1	2001q3	2005q1			
(50) Oklahoma City, OK	2001q2†§					
(51) Olympia, WA	2002q4					
(52) Orlando, FL	2004q4					
(53) Oxnard-Thousand Oaks-Ventura, CA	1998q3	1997q4	2002q3	1997q4	2002q3	2004q4§
(54) Palm Bay-Melbourne-Titusville, FL	2004q2	2001q1	2004q2			
(55) Panama City-Lynn Haven, FL	2003q2	2002q4	2004q1			
(56) Pensacola-Ferry Pass-Brent, FL	2004q2					
(57) Peoria, IL	2001q4§	2001q4§	2005q3			
(58) Phoenix-Mesa-Scottsdale, AZ	2004q4	2003q2	2004q4			
(59) Pittsfield, MA	1998q4					
(60) Portland-Vancouver-Beaverton, OR-WA	2005q1	1997q4§	2004q4			
(61) Port St. Lucie-Fort Pierce, FL	2003q3	2002q2	2003q4			
(62) Prescott, AZ	2004q2	1997q4	2004q2			
(63) Providence-New Bedford-Fall River, RI-MA	2000q2	1998q1	2000q3	1998q1	2000q2	2002q1
(64) Punta Gorda, FL	2002q3					
(65) Redding, CA	2002q3	2000q3	2002q3			
(66) Reno-Sparks, NV	2004q1	2003q1	2004q1			
(67) Riverside-San Bernardino-Ontario, CA	1999q2	1998q1	2002q4			
(68) Sacramento--Arden-Arcade--Roseville, CA	1998q4	1998q1	2000q4	1997q2	2000q1	2004q2
(69) Salem, OR	2005q3	1996q3§	2005q3			
(70) Salinas, CA	1999q1					
(71) San Diego-Carlsbad-San Marcos, CA	1998q1	1997q3	2002q3			
(72) San Francisco-Oakland-Fremont, CA	1997q3	1997q3	2005q4§	1997q2	2000q2	2001q2§
(73) San Jose-Sunnyvale-Santa Clara, CA	1996q3	1999q4	2001q2§	1997q1	2000q1	2001q2§
(74) San Luis Obispo-Paso Robles, CA	1998q4	1997q3	2000q1			
(75) Santa Barbara-Santa Maria-Goleta, CA	2001q3	1997q2	2002q1			
(76) Santa Cruz-Watsonville, CA	1997q1	1997q1	2006q1§	1997q1	2000q2	2001q2§
(77) Santa Rosa-Petaluma, CA	1998q3					
(78) Sarasota-Bradenton-Venice, FL	2004q4	2001q3	2004q4			
(79) Seattle-Tacoma-Bellevue, WA	2004q4					
(80) Spokane, WA	2004q2					
(81) Springfield, MA	1997q4	1997q4	2002q4			
(82) Springfield, OH	2003q4†§					
(83) Stockton, CA	1999q3	1999q1	2004q4	1997q1	1999q4	2004q4
(84) Tallahassee, FL	2005q1	2002q1	2005q1			
(85) Tampa-St. Petersburg-Clearwater, FL	2004q2					
(86) Tucson, AZ	2004q4	2002q4	2004q4			
(87) Tulsa, OK	2002q1§					
(88) Vallejo-Fairfield, CA	1998q2	1998q1	1999q4			
(89) Vero Beach, FL	2004q1					
(90) Visalia-Porterville, CA	2004q2	2001q1	2004q2	1999q1	2002q3	2004q2
(91) Washington-Arlington-Alexandria, DC-VA-MD	2002q4					
(92) Worcester, MA	1998q2	1997q4	2000q2	1997q4	2000q2	2003q4§
(93) Yuba City-Marysville, CA	2002q2	1997q4	2002q2	1997q4	2000q2	2002q3
(94) Yuma, AZ	2004q3					

Note: § negative break point
† not significant at 5% level

Notes: Table presents dates of the breakpoints using the one, two and three-break models. Estimation details shown in section II.

III. Robustness Analysis Results

In this section, we report on various permutations of specifications reported in the paper. Some are referenced in the text of the paper, while others are added for what we hope is completeness. In general, we do not comment on these tables unless their objective is not readily perceived from reading the paper.

III.A. Various Sensitivity Analyses

Online Appendix Table 4: Alternative pre-boom baseline period

Dependent Variable	Price Growth Rate	HMDA Income Growth Rate	HDMA White Share	HDMA Minority Share
	(1)	(2)	(3)	(4)
Relative Quarters = [-14, -11]	0.000	0.010	-0.001	0.009
	(0.005)	(0.009)	(0.005)	(0.006)
Relative Quarters = [-10, -7]	-0.002	0.012	0.007	0.001
	(0.005)	(0.010)	(0.004)	(0.004)
Relative Quarters = [-6, -3]	--	--	--	--
Relative Quarters = [-2, -1]	0.008	0.004	0.010	-0.002
	(0.004)	(0.008)	(0.004)	(0.004)
Relative Quarters = [0, 3]	0.055	0.034	0.010	-0.004
	(0.006)	(0.010)	(0.004)	(0.004)
Relative Quarters = [4, 7]	0.072	0.025	0.006	-0.002
	(0.006)	(0.008)	(0.005)	(0.005)
Relative Quarters = [8, 11]	0.055	0.007	0.001	0.002
	(0.008)	(0.008)	(0.005)	(0.006)
Relative Quarters = [12, 15]	0.037	0.012	-0.003	0.006
	(0.008)	(0.008)	(0.005)	(0.006)
Relative Quarters = [16, 19]	0.024	0.005	-0.005	0.005
	(0.008)	(0.008)	(0.005)	(0.005)
R-squared	0.75	0.31	0.91	0.86
Number of observations	5,846	5,846	6,222	6,222
Dependent variable mean	0.071	0.058	0.782	0.165
Time FEs	X	X	X	X
Area FEs	X	X	X	X

Notes: The table shows points estimates for several dependent variables around the timeline of the housing booms. Models are based on equation (2), with one difference: the baseline period is 3 to 6 quarters prior to the beginning of a housing boom.

Online Appendix Table 5: Income, race and unemployment based on all households

Dependent Variables	BEA Income Growth Rate	School White Share	School Minority Share	Unemployment Rate
	(1)	(2)	(3)	(4)
Relative Year = -2	-0.009	0.002	-0.006	0.002
	(0.003)	(0.004)	(0.004)	(0.002)
Relative Year = -1	-0.004	0.002	-0.004	0.001
	(0.002)	(0.003)	(0.003)	(0.001)
Relative Year = 0	--	--	--	--
Relative Year = 1	-0.001	0.006	-0.003	-0.003
	(0.004)	(0.003)	(0.003)	(0.001)
Relative Year = 2	-0.002	0.005	-0.001	-0.005
	(0.003)	(0.003)	(0.003)	(0.001)
Relative Year = 3	-0.004	0.001	-0.001	-0.005
	(0.003)	(0.003)	(0.003)	(0.001)
Relative Year = 4	-0.000	0.000	-0.002	-0.004
	(0.003)	(0.003)	(0.003)	(0.001)
Relative Year = 5	-0.006	-0.000	-0.000	-0.002
	(0.003)	(0.003)	(0.003)	(0.001)
R-squared	0.451	0.981	0.978	0.859
Number of observations	1,620	1,724	1,724	6,222
Dependent variable mean	0.047	0.610	0.307	0.059
Time FEs	X	X	X	X
Area FEs	X	X	X	X

Notes: The table shows point estimates for many dependent variables based on all households within MSAs, including renters. Models follow equation (2), and all dummies for relative year zero are omitted in the estimation.

Online Appendix Table 6: Refinances and house transactions

Dependent Variables	Log Refinances	Log Transactions
	(1)	(2)
Relative Year = -2	0.065	-0.062
	(0.053)	(0.023)
Relative Year = -1	0.135	-0.031
	(0.040)	(0.021)
Relative Year = 0	--	--
Relative Year = 1	0.234	0.057
	(0.036)	(0.017)
Relative Year = 2	0.326	0.065
	(0.038)	(0.021)
Relative Year = 3	0.306	0.040*
	(0.035)	(0.022)
Relative Year = 4	0.277	-0.005
	(0.037)	(0.021)
Relative Year = 5	0.189	-0.024
	(0.039)	(0.021)
R-squared	0.943	0.973
Number of observations	6,214	6,222
Dependent variable mean	6.39	7.19
Time FEs	X	X
Area FEs	X	X

Notes: The table shows points estimates for log refinances and log number of transactions around the timeline of the housing booms. Models follow equation (2), and all dummies for relative year zero are omitted in the estimation.

Online Appendix Table 7: Growth Rate Variables

Dependent Variables	HDMA White Share Growth Rate	HDMA Minority Share Growth Rate	Prime Share Growth Rate	Subprime Share Growth Rate	FHA Share Growth Rate	Cash-only Share Growth Rate	Average Initial LTV Growth Rate	Variable-rate Share Growth Rate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Relative Year = -2	0.010**	-0.046	-0.016*	0.071	-1.022**	-0.020	0.009*	0.005
	(0.005)	(0.030)	(0.010)	(0.061)	(0.458)	(0.024)	(0.005)	(0.052)
Relative Year = -1	0.013***	-0.037*	-0.017**	0.065	-1.057**	-0.027	-0.000	-0.013
	(0.003)	(0.021)	(0.008)	(0.046)	(0.442)	(0.023)	(0.004)	(0.044)
Relative Year = 0	--	--	--	--	--	--	--	--
Relative Year = 1	0.003	-0.046*	0.007	0.022	-1.212***	-0.046***	-0.002	0.046
	(0.005)	(0.026)	(0.008)	(0.038)	(0.453)	(0.017)	(0.004)	(0.030)
Relative Year = 2	-0.001	0.020	-0.001	0.075**	-1.608***	-0.060***	0.006	0.083
	(0.005)	(0.019)	(0.009)	(0.033)	(0.586)	(0.017)	(0.006)	(0.051)
Relative Year = 3	-0.002	0.019	-0.018*	0.068**	-2.039**	-0.087***	0.013*	0.024
	(0.005)	(0.019)	(0.009)	(0.031)	(0.809)	(0.018)	(0.008)	(0.039)
Relative Year = 4	-0.001	0.044**	-0.023**	0.054*	-2.670**	-0.053**	0.007	0.004
	(0.006)	(0.022)	(0.010)	(0.029)	(1.212)	(0.021)	(0.006)	(0.030)
Relative Year = 5	0.002	0.009	-0.040***	0.027	-3.848**	-0.038	0.002	-0.043
	(0.005)	(0.032)	(0.010)	(0.029)	(1.554)	(0.023)	(0.007)	(0.041)
R-squared	0.185	0.097	0.435	0.326	0.097	0.222	0.311	0.513
Number of observations	5,846	5,822	5,846	5,557	5,046	5,846	5,846	5,560
Dependent variable mean	-0.003	0.04	0.023	0.213	-0.165	0.030	-0.005	0.382
Time FEs	X	X	X	X	X	X	X	X
Area FEs	X	X	X	X	X	X	X	X

Online Appendix Table 8:
Refinance and Sales Volumes

Dependent Variables	Growth in Refinance	Growth in Resale
Relative Year = -2	-0.312	-0.060
	(0.189)	(0.056)
Relative Year = -1	-0.246	-0.063
	(0.167)	(0.063)
Relative Year = 0	--	--
Relative Year = 1	-0.356	-0.117
	(0.242)	(0.103)
Relative Year = 2	0.071	-0.015
	(0.077)	(0.027)
Relative Year = 3	0.048	-0.013
	(0.076)	(0.034)
Relative Year = 4	0.071	-0.047*
	(0.067)	(0.024)
Relative Year = 5	0.016	-0.024
	(0.049)	(0.023)
R-squared	0.395	0.378
Number of observations	5,838	5,846
Dependent variable mean	0.958	0.110
Time FEs	X	X
Area FEs	X	X

Online Appendix Table 9: Speculators

Dependent Variables	Speculator	Speculator Growth Rate
	(1)	(2)
Relative Year = -2	0.012	-0.171
	(0.010)	(0.162)
Relative Year = -1	0.009	-0.149
	(0.008)	(0.095)
Relative Year = 0	--	--
Relative Year = 1	0.008	-0.249**
	(0.008)	(0.110)
Relative Year = 2	0.021***	0.060
	(0.008)	(0.088)
Relative Year = 3	0.018**	0.035
	(0.008)	(0.054)
Relative Year = 4	0.013*	0.025
	(0.007)	(0.040)
Relative Year = 5	0.006	0.016
	(0.006)	(0.040)
R-squared	0.866	0.347
Number of observations	6,222	5,845
Dependent variable mean	0.250	0.271
Time Fes	X	X
Area Fes	X	X

Online Appendix Table 10: LTV > 0.95

Dependent Variables	Share of LTV > 0.95	Share of LTV > 0.95 Growth Rate
	(1)	(2)
Relative Year = -2	2.722***	0.017
	(0.629)	(0.013)
Relative Year = -1	1.929***	-0.017
	(0.518)	(0.010)
Relative Year = 0		
Relative Year = 1	-1.046**	-0.072***
	(0.432)	(0.012)
Relative Year = 2	-2.170***	-0.038**
	(0.572)	(0.016)
Relative Year = 3	-1.704***	0.013
	(0.597)	(0.021)
Relative Year = 4	-1.166*	0.019
	(0.611)	(0.025)
Relative Year = 5	-0.297	0.010
	(0.553)	(0.020)
R-squared	0.828	0.344
Number of observations	6,222	5,846
Dependent variable mean	25.750	-0.010
Time FEs	X	X
Area FEs	X	X

Online Appendix Table 11: Net Migration Robustness Analysis

Dependent Variables	Net Flow of Household Population Growth Rate	Net Flow of Individual Population Growth Rate
	(1)	(2)
Relative Year = -2	-2.787	1.151
	(1.972)	(1.451)
Relative Year = -1	-0.796	0.092
	(0.786)	(0.479)
Relative Year = 0		
Relative Year = 1	-2.265	2.090
	(1.596)	(1.879)
Relative Year = 2	-1.912	3.269
	(1.310)	(2.398)
Relative Year = 3	-1.315	1.651
	(0.969)	(1.460)
Relative Year = 4	-0.472	-1.606
	(0.649)	(2.013)
Relative Year = 5	-0.432	-1.052
	(0.556)	(2.125)
R-squared	0.109	0.090
Number of observations	5,846	5,842
Dependent variable mean	-0.670	0.064
Time Fes	X	X
Area Fes	X	X