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

This Version: February 2, 2018

Abstract

We provide novel estimates of the location, timing, magnitude, and determinants of the start of the last U.S. housing boom. The last housing cycle cannot be interpreted as a single, national event, as different markets began to boom across a decade-long period, some of them multiple times. A fundamental factor, income of prospective buyers, can account for half of the initial jump in price growth, while expansion of purchases by underrepresented minorities cannot. Mortgage financing at the start of the local market booms was conventional, with the share of subprime loans rising sharply only after a multi-year lag.

JEL CODES: R2, R3, E3, and G2.

* Contact information: Fernando Ferreira: email: fferreir@wharton.upenn.edu, address: The Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104. Joseph Gyourko: email: gyourko@wharton.upenn.edu, The Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104. We thank the Research Sponsors Program of the Zell/Lurie Real Estate Center at Wharton for financial support. Qize Chen, Matthew Davis, Anthony DeFusco, Wenjie Ding, Hye Jin Lee, Gordon MacDonald, Xuequan Peng, and Cindy Soo provided outstanding research assistance. We thank Matthew Gentzkow, Ed Glaeser, Erik Hurst, Alvin Murphy, Emily Oster, Paul Willen, and the participants in seminars at the Board of Governors of the Federal Reserve System, the NBER Summer Institute, the University of Chicago, UC-Berkeley, and UCLA for comments on previous versions of the paper. A previous version of this work appeared as the NBER Working Paper #17374 titled "Anatomy of the Beginning of the Housing Boom: U.S. Neighborhoods and Metropolitan Areas, 1993-2009".

I. *Introduction*

The recent U.S. housing cycle was among the defining economic events of the last half century. The magnitudes of the boom and bust as reflected in the Case-Shiller repeat sales home price index was over four times greater than the analogous changes during the previous two cycles. Not surprisingly, much research has been written on the last cycle, most of it focused on the bust and its consequences or on the factors that helped build the boom near its peak.[1]

In contrast, there is virtually no work on when the last housing boom began. This is important for a number of reasons, not least because understanding why it began requires knowledge of when it began. Detailed analysis of the start of a boom also is likely to provide insight into what transpired subsequently. For example, any study of possible spillover effects depends critically on when and how the boom began in specific markets. Whether the boom was fundamentally based at the start helps put in context claims about possible irrational exuberance later in the cycle. How long a time span exists between when a boom starts and the widespread introduction of risky financing structures can help policy makers evaluate the usefulness of different types of regulatory interventions.

In this paper we provide a framework for determining when housing booms begin and estimate it for the most recent cycle. Our dating procedure is consistent with the implications of a dynamic version of a well-known model of spatial equilibrium in urban economics (Glaeser, et. al. (2014)). In that framework, prices in a given market tend to grow at a constant rate along their steady state path, and then discretely jump (fall) if there is a positive (negative) shock to local income or amenities. Since we focus on the start of the boom, the positive jump marks the beginning of the boom. Empirical implementation relies on time series methods (Andrews (1993), Bai and Perron (1998), Bai (1999) and Estrella (2003)), and our estimation is based on finding structural breaks in the house price appreciation time series for individual housing markets. Because a housing market can have more than one boom (and bust) during the sample period, we allow for multiple structural breaks.

[1] The massive amount of work on the housing bust is too voluminous to be cited here. See Ferreira and Gyourko (2015) for a summary on research related to the housing bust itself, and Brunnermeier (2009) and Mian and Sufi (2016) for some of the consequences for consumption decisions and the broader economy. Many factors were proposed to explain the large increase in prices, especially between 2002 and 2005. These include irrational exuberance (Shiller, 2005), subprime mortgages (Mian and Sufi, 2009), expansion of credit via bank deregulation (Favara and Imbs (2015), reductions in mortgage rates (Bhutta and Keys, 2016), and a potential change in the composition of homebuyers (Adelino, Schoar and Severino (2016), Mian and Sufi (2015), Foote and Willen (2016), and Albanesi, De Giorgi, and Nosal (2017)).

We use a large proprietary micro data base from DataQuick with over 23 million housing transactions across 94 metropolitan areas from 1993-2009. Access to rich micro data is critical for our empirical design, as it permits the creation of two random subsamples to separately estimate the timing and magnitude of the beginning of the boom, thereby avoiding the specification search bias that arises when the same data are used to identify both the timing and size of a breakpoint jump (Leamer (1983)).

Our results illuminate key stylized facts about the great housing boom, the first of which is that the boom itself was not a single national event that built smoothly over time. Rather, it was a series of individual market's price growth jumping at specific times across a decade-long period. The first booms started in select MSAs on both coasts before spreading along each coast and then to interior markets. There is also heterogeneity in the number of booms by market, with many MSAs in the Midwest having no booms or just busts, and some markets on both coasts having up to three booms. Another important fact is that the initial price spurts are not inconsequential, averaging about 11% unconditionally, and 5% after controlling for time and market fixed effects. Elevated price growth above pre-boom levels persists for many years.

Knowing the timeline of each market's boom begs the question of what led, jumped coincidentally with, or lagged the initial surge in house prices. Our analysis show that the beginning of the last boom was fundamentally based to a significant extent. The incomes of the pool of potential homebuyers jump simultaneously with, but not before, the escalation in local market price appreciation. Given plausible values of the income elasticity of demand for housing, income growth can account for about 50% of the initial jump in house values. The positive change in fundamentals is short lived though, as the income growth of potential buyers reverts back to its pre-trend in year three of the boom, on average.

Other factors that do not change when housing booms begin are also noteworthy. For example, local market booms did not begin with an expansion of buying by underrepresented minorities. It also is the case that the beginning of the last boom was not associated with a contemporaneous increase in risky loans; nor is there evidence of a prior jump. Rather, the start of local market booms was conventionally financed for the most part. Prime mortgage share increases sharply and significantly in the first year of the boom, while the subprime share increases modestly (but not statistically significantly). However, as the boom builds, subprime shares continue to increase (more than 2% above pre-trends by year 4 of the boom), while prime shares revert to pre-boom levels. Other market data, such as average loan-to-value (LTV) and the share of variable rate loans, are consistent with the subprime mortgage pattern. If the most recent housing boom is typical of the nature of housing cycles, then increasing mortgage and financial risk is a characteristic of a mature boom, not a new one.

Given the amplitude of the last housing cycle, it is not surprising that the government is interested in tracking the evolution of risk in housing markets (e.g., see Fuster, Guttman-Kenney and Haughwout (2016)). The methods developed here further our understanding by providing a framework that exploits the underlying heterogeneity in housing booms – which markets are booming at any point in time, their timelines, and price changes - and that more precisely links the variation from each location with fundamental or credit market factors. Moreover, the timeline of local housing booms may also be useful in understanding the interaction of housing cycles with other aspects of the economy, such as the educational achievement of individuals (Charles, Hurst and Notowidigdo, 2016) or the determinants of local public finances (Davis and Ferreira, 2017).

The remainder of the paper proceeds as follows. The next section outlines the conceptual framework and empirical strategy for identifying the beginning of a housing boom. Section III follows with a description of our data sources. Results on the timing, magnitudes, and potential determinants of the beginning of the housing booms are presented in Section IV. Section V concludes.

II. Theory and empirical methods

Our empirical strategy involves using structural breaks in the time series of local housing price appreciation rate series to pin down the start of a housing boom. A motivation for this approach arises from the core model of spatial equilibrium in urban economics introduced by Rosen (1979) and Roback (1980, 1982). Their compensating differential framework models the location decision as a function of utility based on the wages (W_i) one earns by working in labor market area *i* (wages are a function of the area's productivity), the amenities (A_i) consumed in the same market, and the house price (H_i) one has to pay to live there. Equalizing utility implies that housing prices are the entry fee a household must pay to access the productivity and amenities of a local labor market.

A highly simplified, linearized version of the model would posit utility as the sum of local wages and amenities less local house value, as in equation (1):

(1) $U^* = W_i + A_i - H_i$.

Differentiating yields

(2) $dU^* = dW_i + dA_i - dH_i = 0$ or $dW_i + dA_i = dH_i$.

Equation (2) illustrates that higher house prices will exist in equilibrium in markets with high wages (productivity) and valuable amenities.

More relevant for our purposes is the dynamic framework introduced in Glaeser, et. al. (2014). Their framework includes two difference equations, one for the supply of housing and one for demand that fully describes the housing market. Risk-neutral builders are presumed to profit maximize and compete so that they just cover their costs while earning a normal entrepreneurial return. Thus, expected house values equal the full costs of production. Housing demand is determined in equilibrium by equating the net flow costs of living in the reservation market that always provides the minimum utility level to the net benefits of living there over the same time period.

Those equations fully describe housing supply and demand when combined with a transversality condition preventing bubbles (see Glaeser, et. al. (2014) for the details). Together, they generate steady state values for house prices, new construction, and city size for each market. If there are no shocks to any local traits (i.e., $dW_{i,t}=0$ and $dA_{i,t}=0$ for all t) and city size equals its optimal value N* for some initial period, then those steady state values fully describe location i. In that case, the market will have constant growth path values for house prices, new construction and population.

House price appreciation should then exhibit a discrete jump if there is a positive shock to local productivity or amenities. That jump marks the beginning of the boom. More formally, this suggests we should search for structural breaks in the price growth series as markers for when cycles begin (and end).

Consider the following reduced form model of house price growth in MSA i at time t:

(3) $PG_{i,t} = d_{i,t} + \epsilon_{i,t}, t = 1, \dots, T.$

The discussion above implies that $d_{i,t} = d_{i,0}$ for all t if the market is on its steady-state growth path. However, if there is a positive shock at time t then the price growth rate will exhibit a discrete jump in that period, and $d_{i,t} \neq d_{i,0}$. To carry out this test, we follow established methods in the time series literature for estimating such breaks. The null hypothesis is that $d_{i,t}$ is constant for the entire sample period:

$$H_0: d_{i,t} = d_{i,0}, t = 1, \dots T.$$

The alternative is that $d_{i,t}$ changes at some proportion, $0 < \pi_i < 1$, of the sample which marks the beginning of a housing boom in market *i*. The alternative hypothesis is

$$H_1:\ d_{i,t} = \begin{cases} d_{1,i}(\pi_i), t = 1, \dots, \pi_i T \\ d_{2,i}(\pi_i),\ \ t = \pi_i T + 1, \dots T. \end{cases}$$

Let $\Pi_{\mathrm{i}} = [\pi_{i,1}, \pi_{i,2}]$ be a closed interval in $(0,1)$ and let S_i be the set of all observations from $t = int(\pi_{i,1}T)$ to $t = int(\pi_{i,2}T)$, where $int(\cdot)$ denotes rounding to the nearest integer. The estimated breakpoint is the value t^* from the set S_i that maximizes the likelihood ratio statistic from a test of H_1 against H_0. Direct calculation of the probabilities in the likelihood ratio is non-trivial and prior research has relied on approximations that typically are based on simulation or curve-fitting methods (Andrews (1993), Hansen (1997)). However, Estrella (2003) provides a numerical procedure for calculating exact p-values that does not rely on these types of approximations. We use this method to calculate p-values for the estimated break point, π_i, for each MSA in the sample.

Our procedure will generate a breakpoint estimate regardless of whether the structural break represents a positive or negative change in the price growth rate. When an estimated breakpoint is insignificant, we conclude that the market did not have a boom. When we find a statistically significant breakpoint, we also test for the existence of two breaks against the null hypothesis of only one, closely following Bai (1999) and Bai and Perron (1998). If we can reject the null hypothesis of one break against the alternative of two, we discard the one-break model estimate and keep the results from the two-break model. Similarly, for those MSAs with two breaks, we also estimate and test for the significance of three breaks relative to two. Allowing for multiple breaks helps address the possibility that certain MSAs had more than one shock or cycle during the sample period. We also tested a Markov switching model that estimates probabilities of being in different regimes (Hamilton (2016)). Those probabilities largely line up with our estimated breakpoints, but we prefer using the breakpoint methodology because it provides sharp estimates that allow us to see what other forces might have changed along a well-defined time line of the boom.

Note that the breakpoint method does not provide an unbiased estimate of the magnitude of the change in price growth rates at the breakpoint. Under the null hypothesis that there is no breakpoint, the estimate of $d_{i,t}$ has a nonstandard distribution and OLS estimates of its magnitude will be upwardly biased in absolute value. This can lead to an increased chance of falsely concluding a break exists, and is a form of specification search bias arising from the fact that the same data is being used to estimate both the timing and the magnitude of the structural break (Leamer (1983)). Our approach to correcting the magnitude estimates follows Card, Mas, and Rothstein (2008) in randomly splitting the underlying sample of houses transactions into two, one for estimating breakpoints and the other to estimate the magnitude of price changes. If the two subsamples are independent, then estimates of $d_{i,t}$ from the second sample, which was not used to estimate the location of the breakpoint, will have a standard distribution even under the null hypothesis of no structural break in the first sample.

The holdout panel includes all markets (i) and quarters (t) available. Magnitudes of the changes in price growth rates (*PG*) are estimated for all pre- and post-boom years, based on the following panel equation:

$$(4)\quad PG_{i,t} = \sum_{\rho=-6}^{6} \sum_{b \in B} \theta_\rho^b \, \mathbf{1}\{t - t^{*,b} = \rho\} + q_t + a_i + \epsilon_{i,t}$$

where $\mathbf{1}\{t - t^{*,b} = \rho\}$ is a set of years relative to the quarter of the beginning of the housing boom ($t^{*,b}$). For example, relative year 1 includes the quarter of the breakpoint plus three subsequent quarters. Relative year 2 dummy includes the next four quarters (#4-#7 since the boom started), and so on. We show all estimates up to relative year five, as after that there is a large drop in the number of MSAs with relevant data (all models also include dummies for relative years 6 or above). To distinguish between markets that had a housing boom and those that did not, we do this separately for breakpoints that are positive and statistically significant

($b = boom$), and those that are either negative or insignificant. The coefficients, θ_ρ^{boom}, then describe how prices in a market that had a boom evolve over the course of time. These estimates are relative to the 12-month period prior to the beginning of the boom (the dummy for that period is excluded from the estimation). Additionally, q_t are year-quarter fixed effects, and a_i are MSA fixed effects.

III. Data

Our primary source of housing market data comes from DataQuick (now CoreLogic). This firm provides micro observations on home purchases collected from deeds records. The final sample of DataQuick observations used in this analysis contains more than 23 million arms-length, single-family and condominium housing transactions in 94 metropolitan areas spread across 29 states, from 1993(Q1) to 2009(Q4). Transactions are included only if they are in MSAs that have consistent data at least since 1998(Q1) and if they are from MSAs with full or nearly full coverage of their component counties. The first quarter of 1998 was chosen as the latest possible starting point to help ensure that we have enough information to estimate the beginning of the boom. Online Appendix Table 1 reports the representativeness of our final sample.

We observe the exact date, price, and location of each housing transaction, in addition to a variety of housing traits. Using these data we create a MSA-level constant quality house price series by quarter using a hedonic model.[2] Transaction price, in logarithmic form, is modeled as a function of the square footage of the home entered in quadratic form, the number of bedrooms, the number of bathrooms, and the age of the home. We also include a dummy for condominiums or houses located in subdivisions and interact these dummies with the linear and quadratic terms for square footage. The hedonic index values are derived from the estimated coefficients on the year-quarter dummies, and then normalized to 100 in 2000(Q1) for all MSAs. Figure 1 plots each MSA price index, and it is obvious that dispersion in prices changed dramatically over the cycle.

DataQuick does not provide demographic information about individual homeowners, so we merge their housing transactions with the Home Mortgage Disclosure Act (HMDA) files in

[2] We employ hedonic price indexes because their data requirements are less onerous than repeat sales indexes, but we have investigated both methodologies and find they yield very similar results.

order to capture information on the income and race of all loan applicants.[3] These data include self-reported information for all mortgage applicants, not just those whose loans are approved and go on to buy a house. About one-third of the loans in HMDA are rejected, withdrawn, or approved but do not result in a purchase, so our demographic measures are based on a pool considerably larger than just those who bought homes. It also has quarterly variation, so it can be precisely calculated around the breakpoint.

We also look at labor market conditions and demographics for all residents in the metropolitan area, including all renters and current homeowners that are not necessarily planning to buy a new house. The local unemployment rate is measured monthly by the Bureau of Labor Statistics (BLS); we use quarterly averages in our analysis. From the Bureau of Economic Analysis (BEA), we collect yearly data on overall personal income by MSA. And, from the Common Core of Data (CCD), we collect the yearly demographic composition of all students in each metropolitan area.

DataQuick also provides information on the amounts of up to three loans used to finance the purchase of each home, as well as the names of the buyers, sellers, and lender(s). There is no one formal or legal definition of what constitutes a subprime loan. Researchers use either FICO scores or lender lists to define what constitutes a subprime loan. We follow the latter procedure. More specifically, we obtained lists of the top twenty subprime lenders from 1990-onward in a publication now called *Inside Mortgage Finance* (this publication claims to capture up to 85% of all subprime originations in most years) and combined it with the complete HUD list of subprime lenders.[4] Our subprime measure is the share of all purchasers with current mortgages issued by these lenders.

To help insure we do not conflate subprime and prime loans, we separately identify borrowers whose loans were guaranteed by FHA or VA (regardless of lender identity). 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. We also separately categorized purchasers who were financed by individuals, households or firms that

[3] Both data sets were then merged using a straightforward sequential matching process. In total, 93.2% of all sales transactions in DataQuick were matched at some point in the procedure. Of those, approximately 60% can be considered "high quality" matches, as there was a unique match based on tract ID, year of transaction, precise house value, and lender name.

[4] See Chomsisengphet & Pennington-Cross (2006) for more detail on these lenders and lists.

issued less than 100 loans throughout our sample period. All remaining owners with debt are Prime borrowers by definition. Homes purchased without a mortgage are classified as cash-only.

Finally, we measure two other proxies for riskier loans: i) the average loan-to-value ratio of all transactions in a quarter, which is based on sum of the mortgage balance of up to three initial loans per property divided by the transaction price, and ii) the share of home purchase mortgages in a quarter that have some type of variable interest rate structure.[5] Table 1 reports descriptive statistics on each of these variables.

IV. Results

a. Timing and magnitude of local booms

Figure 2 plots the histogram of the timing of breaks for all statistically significant positive structural breaks. This shows that the last boom is much more accurately characterized as a series of local booms that ran from the mid-1990s to the mid-2000s, rather than as a single national event. Most quarters see at least one market start to boom, and many have from two to four percent of the total number of breaks. The exception is the 2nd quarter of 2004, with about 10% of all breaks. There are no MSAs that started a housing boom after the third quarter of 2005.

More detail on breakpoints in all markets is reported in Online Appendix Table 2. Those data show there were not any statistically significant jumps in the rate of price growth in six markets, with most being in the Midwest or upstate New York. Negative and statistically significant jumps were found in 24 instances. Some of these places, such as Cleveland and Detroit, are in long-term decline and never experienced a meaningful housing boom during the time span of our data. Eighty-two markets had at least one statistically significant boom, with starting dates ranging widely in time from the last quarter of 1995 in Boston, MA, to the third quarter of 2005 in Salem, OR, and Mt. Vernon-Anacortes, WA. Fifty-eight markets had at least two jointly statistically meaningful jumps in price growth, and in 21 markets we estimated three jointly significant structural breaks. They are usually coastal markets that had an early boom prior to 2000 and a second boom in the mid-2000s. The three-break model estimates are also able to capture some negative breaks from areas that had house price declines early in the 2000s.

---

[5] Interest rates themselves are not used in our empirical analysis given that our empirical strategy is best suited to examining factors that vary across space and time; mortgage rates are roughly constant across local markets.

Table 2 summarizes the results of estimating equation (4), showing changes from two years prior to the start of the boom to five years after it. The first column reports the magnitude of the jump for any significantly positive estimate when no time or location fixed effects are controlled for. Price growth is 10.7% higher in relative year 1. Column 2's numbers are from a specification including time dummies, so they represent the variation remaining after sweeping out the impacts from nationwide changes that are common across all locations. The estimated jump at the beginning of the typical boom is now 4.5%, so over 50% of the variation in unconditional price changes when booms start can be explained by factors that are not exclusively local. Even though the timing of the start of local booms does not seem to follow a national pattern, yearly price changes do. Column 3's results are from our preferred specification that also includes area fixed effects. There is little change from column 2: price growth rates jump by 5.1% at the beginning of the boom relative to the pre-boom year, then go to 6.8%, 5.1%, 3.4%, and 2.2% in subsequent years. Thus, it takes at least five years on average after the start of a boom to get back to the price growth rate that existed before a significantly positive break.[6]

Columns 4 and 5 also report estimates for significantly negative breaks (i.e., busts) and for statistically insignificant breaks. For those market experiencing busts, price growth in relative year 1 is very negative and remains so for many years afterwards. Not unexpectedly, prices were mostly flat in markets without a statistically meaningful jump in either direction. In the remainder of the paper we focus solely on the timeline of the booms.

b. Incomes and Minority Status

Table 3 reports results regarding how other factors changed across the timelines of local housing booms. Income is a fundamental in any model of housing demand, so the first column of Table 3 reports results for the income of potential home buyers. It jumps by a statistically meaningful 3.3% at the beginning of the local area's housing boom and stays elevated

[6] These results show a modest pre-trend, with a small, but statistically significant, negative growth rate in relative year -1. Further analysis suggests this likely is an artifact of measurement error in the estimation of breakpoints, as the breaks have a non-zero probability of being one quarter earlier or later than the dates shown in Table 2. To investigate this more fully, we estimated another specification that added a dummy for relative quarters -1 and -2. In that specification, the baseline period becomes relative quarters -3 to -6, which are less likely to be affected by measurement error. The results, which are reported in the first column of Online Appendix Table 3, show the complete absence of a pre-trend.

throughout the following year, before reverting back to pre-boom rates of growth. In addition, there is no evidence of a pre-trend in the income of those applying for a mortgage this year, as the coefficients for relative years -1 and -2 are much smaller and not significantly different from zero. Hence, this variable does not jump significantly prior to the boom in house price growth.

To help gauge the economic significance of this result, we can appeal to the housing economics literature which suggests an income elasticity of demand that ranges from 0.75 to 1.[7] Given these numbers, the jump in income can account for around half of the initial jump in prices at the beginning of the boom. While the beginning of a boom is at least partially fundamentally grounded, we caution against concluding the same for the overall cycle. The spike in income in relative year 2 is less than half that for price growth, and income growth reverts back to pre-boom levels from relative year 3-onward. That was not the case for the price appreciation path, so there is no reasonable income elasticity of demand that can explain very much of the additional growth in house prices as the boom builds.

Columns 2-3 of Table 3 report results on the variation in the racial composition of buyers across the housing cycle timeline. There is a 0.8 percentage point change in White buyer share in the year a boom begins. This does not persist in a significant fashion for another year and the White share has reverted back to pre-boom levels entirely by relative year 3. Column 3 reports a drop in minority share as the boom begins, but this is a small and insignificant change. There is no statistically or economically meaningful shift in minority buyer shares at any point along the timeline of the boom.[8]

c. Mortgages

---

[7] See Polinsky and Ellwood (1979) for a classic early examination that puts the elasticity in the below 1 range. Subsequent research that generates slightly higher estimates claims that the larger impact is due to the use of better measures of permanent income (e.g., Goodman and Kawai (1982); see also Quigley (1979)).

[8] For completeness, we also estimated magnitudes for incomes and race using different measures based on all households in a market, not just those mostly likely to be on the margin for buying a home. Those results are reported in Online Appendix Table 4. These estimates do not show a significantly positive jump in relative year 1 or 2. This is likely explained by measurement error biasing the coefficients towards zero. They are annual measures (not quarterly), and they reflect a much larger population, some of whom are renters not necessarily interested in becoming an owner-occupier in any given year, while others are existing owners who may not be mobile. These results highlight the importance of having measures that pertain as closely as possible to the pool of marginal buyers. This appendix table also shows how overall unemployment rates varied over the cycle. That rate declines by a statistically significant 0.3 percentage points in the first year of the boom (compared to the pre-boom year), and by 0.5 percentage points in relative years 2 and 3, before slowly starting to return to pre-boom levels in subsequent years.

Columns 4 to 9 of Table 3 document how the shares of purchasers using different types of mortgages vary over the timeline of housing booms. Note that there is no evidence of a meaningful jump in prime or subprime share in the years prior to the start of a local market boom, but this is not the case for FHA-insured borrowers. Thus, there is no evidence that private lenders (whether conventional or subprime) outside the FHA system anticipate the start of housing booms and increase mortgage lending just before housing prices start to jump.

In relative year 1, we then see a sharp jump of 1.2 points in Prime borrower share, and a much smaller (and not statistically significant) 0.4 point higher Subprime borrower share. Lending starts to ramp up only when prices and incomes jump, and these findings indicate that the bulk of the increase is being taken from the FHA-insured sector. This is consistent with the concomitant rise in the income of current homebuyers, as households with rising incomes reasonably could be viewed as less risky, so that they would be more likely to use the prime mortgage market rather than the more expensive FHA market.

The prime borrower share grows by an even larger 1.9 points each of the following two years. It does not revert back to pre-boom levels until relative year 5. Subprime share does finally jump significantly in relative year 2—by 1.1 percentage points. And, it keeps going up after that, and does not mean revert back to pre-boom levels throughout our timeline.[9] Finally, the cash-only buyer share is flat in the first and second years of the typical boom, but it does start to drop in a statistically significant fashion three years into the boom.

This increase in riskier lending with a lag from the start of a boom is corroborated by other proxies for risky mortgages reported in the final two columns of Table 3. LTVs at purchase have very small and insignificant positive point estimates in the first two years after the beginning the boom, and then increase by an economically and statistically robust 1.3, 1.5, and 1.5 percentage points, respectively, in relative years 3, 4 and 5. The share of adjustable rate mortgages follows a similar post-boom pattern.[10]

[9] This is consistent with the conclusion in Brueckner, Calem and Nakamura (2011) that subprime lending is a consequence, rather than a cause, of bubble-like conditions in housing markets.

[10] In Online Appendix Table 5 we also show magnitude estimate for volume of transactions and refinances. We find a statistically significant jump in transactions volume when the boom first begins, and that high volume is maintained for the next couple of years, too. Refinancings are higher than baseline in almost all years, and become especially large in relative years 2 and 3, coinciding with the peak of price growth rates. Pre-trend estimates are noisier for both volume variables, but somewhat consistent with DeFusco, Nathanson and Zwick (2017) who document a lead-lag relationship between changes in trading volume and price changes in U.S. housing markets.

V. *Conclusion*

A glaring omission in the recent literature investigating the causes and consequences of the last housing cycle is the absence of estimates of the beginning of a housing boom. This is not surprising given the difficulties in conceptualizing and estimating the beginning of a boom. Guided by the intuition of a dynamic model of spatial equilibrium across housing markets, we address both issues and provide the first timeline of the last boom based on breakpoints in local market price growth rate series.

Our estimates reveal a number of new stylized facts. National aggregate prices mask substantial heterogeneity across and within markets. Housing booms began in different markets as early as the mid-1990s and as late as the mid-2000s; the typical market experienced multiple booms. Thus, the last boom was not a homogenous event that started everywhere at roughly the same time. On average and after controlling for location and time fixed effects, the rate of house price appreciation jumped by about 5 percentage points at the beginning of the typical local market boom.

When we investigate how income, demographic and mortgage conditions vary across the timelines of local housing market booms, there is no evidence that any of these sectors cause housing booms in the Granger sense, as none lead the start of booms. The income of the pool of prospective homebuyers does jump coincidentally with the initial increase in the rate of house price appreciation, so the start of housing booms is closely linked to an important demand-side fundamental. This does not mean that the full cycle is fundamentally grounded, as the increase in prices in subsequent years is not matched by similar increases in homebuyer income. The racial composition of buyers did not change much at the beginning of booms. Finally, the financing of the beginning of booms is conventional for the most part. The share of risky loans, as proxied for by subprime borrower share, increases sharply over time, but only with a lag to the start of the boom.

References

Adelino, Manuel, Antoinette Schoar, and Felipe Severino. 2016. "Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class." *The Review of Financial Studies* 29(7): 1635-70.

Albanesi, Stefania, Giacomo De Giorgi, and Jarmomir Nosal. 2017. "Credit Growth and the Financial Crisis: A New Narrative." National Bureau of Economic Research Working Paper 23740.

Andrews, Donald WK. 1993. "Tests for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica* 61(4): 821-56.

Bai, Jushan. 1999. "Likelihood ratio tests for multiple structural changes." *Journal of Econometrics* 91(2): 299-323.

Bai, Jushan, and Pierre Perron. 1998. "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica* 66(1): 47-78.

Bhutta, Neil, and Benjamin J. Keys. 2016. "Interest Rates and Equity Extraction during the Housing Boom." *American Economic Review* 106(7): 1742-74.

Brueckner, Jan, Paul Calem and Leonard Nakamura. 2012. "Subprime Mortgages and the Housing Bubble." *Journal of Urban Economics* 71(2): 230-243.

Brunnermeier, Markus. 2009. "Deciphering the Liquidity and Credit Crunch 2007-2008." *Journal of Economic Perspectives* 23(1): 77-100.

Card, David, Alexandre Mas and Jesse Rothstein. 2008. "Tipping and the Dynamics of Segregation." *Quarterly Journal of Economics* 123(1): 177-218.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2016. "The Masking of the Decline in Manufacturing Employment by the Housing Bubble." *Journal of Economic Perspectives* 30(2): 179-200

Chomsisengphet, Souphala and Anthony Pennington-Cross. 2006. "The Evolution of the Subprime Mortgage Market." *Federal Reserve Bank of St. Louis Review* 88(1): 31-56.

Davis, Matthew and Fernando Ferreira. 2017. "Housing Disease and Public School Finance." National Bureau of Economic Research Working Paper 24140.

DeFusco, Anthony A., Charles G. Nathanson, and Eric Zwick. 2017. "Speculative Dynamics of Prices and Volume." National Bureau of Economic Research Working Paper 23449.

Estrella, Arturo. 2003. "Critical Values and P Values of Bessel Process Distributions: Computation and Application to Structural Break Tests." *Econometric Theory* 19(6): 1128-43.

Favara, Giovanni, and Jean Imbs. 2015. "Credit Supply and the Price of Housing." *The American Economic Review* 105(3): 958-92.

Ferreira, Fernando, and Joseph Gyourko. 2015. "A New Look at the U.S. Foreclosure Crisis: Panel Data Evidence of Prime and Subprime Borrowers from 1997 to 2012". National Bureau of Economic Research Working Paper 21261.

Foote, Christopher, and Paul Willen. 2016. "Cross-Sectional Patterns of Mortgage Debt during the Housing Boom: Evidence and Implications." Federal Reserve Bank of Boston Research Department Working Papers 16-12.

Fuster, Andreas, Benedict Guttman-Kenney, and Andrew Haughwout. 2016. "Tracking and stress-testing US household leverage." Federal Reserve Bank of New York Staff Report 787.

Glaeser, Edward, Joseph Gyourko, Eduardo Morales and Charles Nathanson. 2014. "Housing Dynamics: An Urban Approach." *Journal of Urban Economics* 81: 45-56.

Goodman, Allen and Masahiro Kawai.1982. "Permanent Income, Hedonic Prices and Demand for Housing: New Evidence." *Journal of Urban Economics* 12(2): 214-237.

Hamilton, James D. 2016. "Macroeconomic Regimes and Regime Shifts." In *Handbook of Macroeconomics*, Vol 2, edited by John B. Taylor and Harald Uhlig, 163-201. Amsterdam: North Holland.

Hansen, Bruce. 1997. "Approximate Asymptotic P Values for Structural-Change Tests." *Journal of Business & Economic Statistics* 15(1): 60-67.

Leamer, Edward. 1983. "Let's Take the Con Out of Econometrics." *American Economic Review* 73(1): 31-43.

Mian, Atif and Amir Sufi. 2009. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *Quarterly Journal of Economics* 124(4): 1449-96.

____________________. 2015. "Household Debt and Defaults from 2000 to 2010: Facts from Credit Bureau Data." National Bureau of Economic Research Working Paper 21203.

____________________. 2016. "Who Bears the Cost of Recessions? The Role of House Prices and Household Debt." In *Handbook of Macroeconomics*, Vol 2, edited by John B. Taylor and Harald Uhlig, 255-96. Amsterdam: North Holland.

Polinsky, A Mitchell and Ellwood, David T. 1979. "An Empirical Reconciliation of Micro and Grouped Estimates of the Demand for Housing." *The Review of Economics and Statistics* 61(2): 199-205.

Quigley, John. 1979 "What Have We Learned About Urban Housing Markets?" in *Current Issues in Urban Economics* (P. Mieszkowski and M. Straszheim, eds.), The Johns Hopkins University Press, Baltimore, MD.

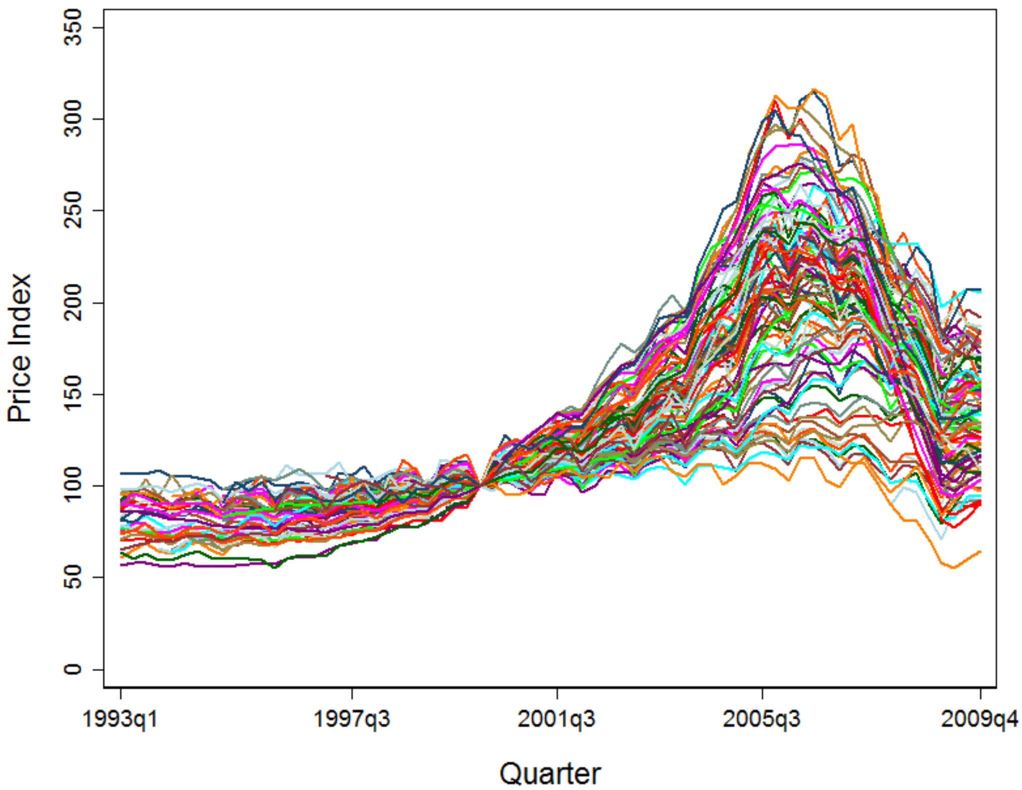
Roback, Jennifer. 1980. "The Price of Local Amenities: Theory and Measurement." Economic Growth Center, Yale University, Center Discussion Paper 359.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90(4): 1257-78.

Rosen, Sherwin. 1979. "Wage-Based Indexes of Urban Quality of Life." In *Current Issues in Urban Economics*, edited by Peter Mieszkowski and Mahlon Straszheim. Baltimore: Johns Hopkins University Press.

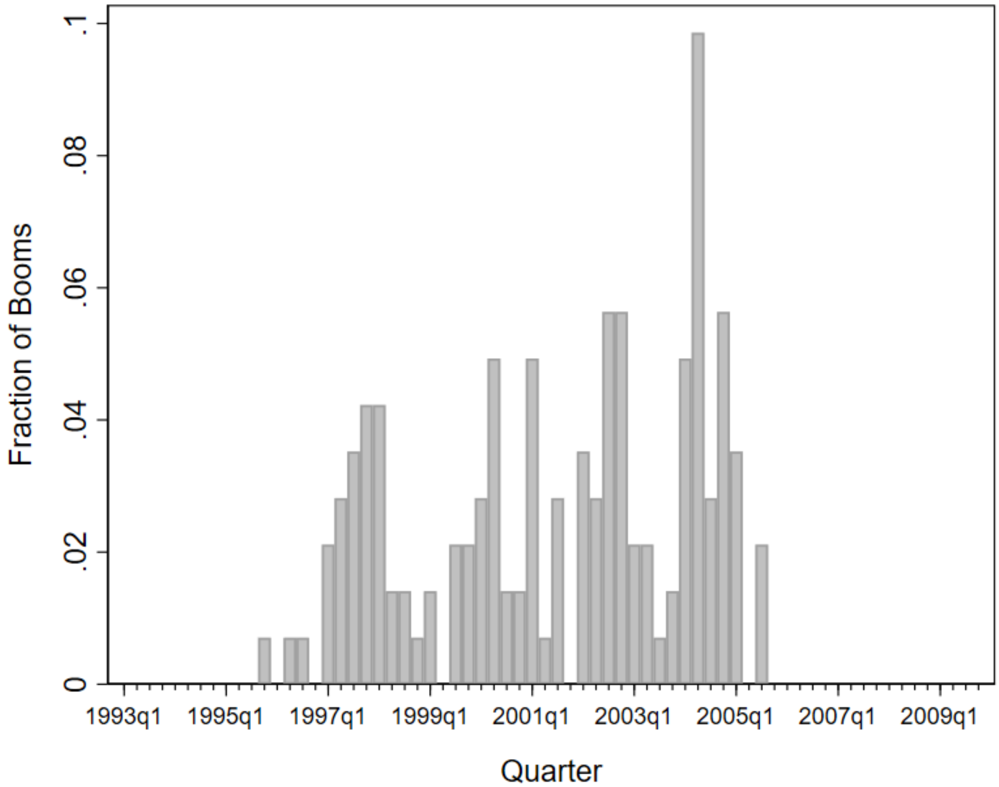
Shiller, Robert. 2005. *Irrational Exuberance*. 2ed. Princeton: Princeton University.

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.

Figure 2. Beginning of local booms histogram



Notes: The histogram plots the fraction of booms that start in a given quarter.

Table 1: Descriptive statistics

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.

Table 2: Magnitude of house price appreciation changes

Dependent Variable	Price Growth Rate				
	Positive Breaks	Positive Breaks	Positive Breaks	Negative Breaks	Non-Significant Breaks
	(1)	(2)	(3)	(4)	(5)
Relative Year = -2	-0.009	-0.022	-0.006	-0.021	0.002
	(0.005)	(0.004)	(0.005)	(0.011)	(0.018)
Relative Year = -1	-0.001	-0.023	-0.012	-0.004	0.009
	(0.005)	(0.003)	(0.004)	(0.010)	(0.015)
Relative Year = 0	--	--	--	--	--
Relative Year = 1	0.107	0.045	0.051	-0.084	-0.000
	(0.007)	(0.004)	(0.005)	(0.016)	(0.030)
Relative Year = 2	0.131	0.061	0.068	-0.076	0.042
	(0.007)	(0.005)	(0.006)	(0.015)	(0.045)
Relative Year = 3	0.096	0.045	0.051	-0.098	0.004
	(0.007)	(0.006)	(0.008)	(0.017)	(0.036)
Relative Year = 4	0.046	0.030	0.034	-0.089	-0.013
	(0.010)	(0.007)	(0.008)	(0.019)	(0.026)
Relative Year = 5	-0.011	0.018	0.022	-0.049	0.016
	(0.011)	(0.006)	(0.008)	(0.019)	(0.028)
R-squared	0.36	0.73	0.75	0.75	0.75
Number of observations	5,846	5,846	5,846	5,846	5,846
Dependent variable mean	0.082	0.082	0.082	0.082	0.082
Time Fes		X	X	X	X
Area FEs			X	X	X

Notes: The table shows points estimates for price growth rates around the timeline of the housing booms and busts. All dummies for relative year zero are omitted in the estimation. Columns 1 and 2 report point estimates for different models based on equation (4). Columns 3, 4, and 5 are estimates from one model that also follows equation (4).

Table 3: Income, race, and financial factors around the timeline of housing booms

Dependent Variables	HMDA Income Growth Rate	HMDA White Share	HMDA Minority Share	Prime Share	Subprime Share	FHA Share	Cash-only Share	Average Initial LTV	Variable-rate Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Relative Year = -2	0.009	0.002	0.006	0.007	-0.005	0.020	-0.013	0.020	0.013
	(0.009)	(0.005)	(0.005)	(0.008)	(0.004)	(0.007)	(0.006)	(0.006)	(0.007)
Relative Year = -1	0.007	0.008	-0.000	0.001	-0.001	0.014	-0.009	0.015	0.003
	(0.009)	(0.004)	(0.004)	(0.006)	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)
Relative Year = 0	--	--	--	--	--	--	--	--	--
Relative Year = 1	0.033	0.008	-0.003	0.012	0.004	-0.007	-0.002	0.002	0.007
	(0.010)	(0.004)	(0.004)	(0.005)	(0.003)	(0.005)	(0.005)	(0.005)	(0.007)
Relative Year = 2	0.025	0.005	-0.001	0.019	0.011	-0.019	-0.005	0.005	0.015
	(0.008)	(0.005)	(0.005)	(0.006)	(0.004)	(0.005)	(0.006)	(0.006)	(0.008)
Relative Year = 3	0.007	0.000	0.003	0.019	0.018	-0.019	-0.013	0.013	0.027
	(0.008)	(0.005)	(0.006)	(0.006)	(0.004)	(0.005)	(0.006)	(0.006)	(0.010)
Relative Year = 4	0.012	-0.004	0.006	0.012	0.024	-0.013	-0.018	0.015	0.029
	(0.009)	(0.005)	(0.006)	(0.006)	(0.004)	(0.006)	(0.005)	(0.005)	(0.011)
Relative Year = 5	0.005	-0.006	0.005	-0.003	0.026	-0.004	-0.017	0.015	0.038
	(0.008)	(0.005)	(0.005)	(0.006)	(0.004)	(0.007)	(0.005)	(0.005)	(0.012)
R-squared	0.310	0.910	0.861	0.837	0.792	0.823	0.818	0.817	0.819
Number of observations	5,846	6,222	6,222	6,222	6,222	6,222	6,222	6,222	6,222
Dependent variable mean	0.062	0.782	0.162	0.531	0.124	0.133	0.176	0.696	0.177
Time FEs	X	X	X	X	X	X	X	X	X
Area FEs	X	X	X	X	X	X	X	X	X

Notes: The table shows points estimates for income, race, and financial variables around the timeline of the housing booms. Each column reports estimates of a separate model based on equation (4). Dummies for relative year zero are omitted in the estimation.