Anatomy of the Beginning of the Housing Boom:
U.S. Neighborhoods and Metropolitan Areas, 1993-2009[*]

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

Abstract

We provide novel estimates of the timing, magnitudes, and potential determinants of the start of the last housing boom across American neighborhoods and metropolitan areas (MSAs) using a rich new micro data set containing 23 million housing transactions in 94 metropolitan areas between 1993 and 2009. We also match transactions data with loan information, enabling us to observe household income and other demographics for each neighborhood. Five major findings are reported. First, the start of the boom was not a single, national event. Booms, which are defined by the global breakpoint in an area's price appreciation series, begin at different times over a decade-long period from 1995-2006. Second, the magnitude of the initial jump in house price appreciation at the start of the boom is economically, not just statistically, significant. On average, log house prices are over four points higher during the first year of the boom relative to the previous twelve month period for both MSAs and neighborhoods. There is no evidence that price growth was trending up prior to the start of the boom. Third, local income is the only potential demand shifter found that also had an economically and statistically significant change around the time that local housing booms began. Contemporaneous local income growth is large enough to account for half or more of the initial jump in house price appreciation. Fourth, there is important heterogeneity in that result. Income growth is large and jumps at the same time as house price appreciation in areas that boomed early and have inelastic supplies of housing, but not in late booming areas and those with elastic supply sides. While these estimates indicate that the beginning of the boom was fundamentally justified on average, they do not imply that what followed was rational. Fifth and finally, none of the demand-shifters analyzed show positive pre-trends, but some such as the share of subprime lending, do lag the beginning of the boom. This suggests that key players in the lending market more responded to the boom, rather than caused it to start.

I.   Introduction

The United States recently experienced house price growth of unprecedented scale, as documented in Figure 1's plot of the national S&P/Case-Shiller and Federal Housing Finance Administration (FHFA) real price indexes. Many researchers have tried to understand whether the most recent cycle was a bubble, or if rational theories can account for the variation in prices and quantities at the national level and across metropolitan areas (MSAs).[1] Despite this work and the fact that we are now several years into the current housing crisis, researchers and policy makers still have conflicting views and limited knowledge about the causes of that extraordinary rise and decline in house prices.[2] This may be explained in part by difficulties of estimating where, when, and how local housing booms start. Theoretically, the housing economics literature is agnostic about how to define the beginning of a boom. Empirically, there has not been publicly available micro data with high quality information about both prices and fundamentals, including the necessary coverage by time and geography for the analysis.[3]

Establishing a time line for the housing boom is important for a variety of reasons. To understand why it started requires knowing when it began. Detailed analysis of the beginning of the boom is also likely to inform our knowledge of what transpired subsequently. For example, any study of contagion effects depends critically on when and how the boom began in specific areas. Moreover, the time line would help improve our understanding of the bust, as many factors correlated with the current foreclosure crisis could also be responding to the boom instead of causing it.

---

[1] Shiller (Chapter 2, 2005) provides perhaps the most famous characterization of the boom as a non-rational event, utilizing both aggregate and metropolitan area data. Others recently have estimated rational expectations general equilibrium models to try to explain the national aggregate price data (e.g., Favilukis, Ludvigson, and Van Nieuwerburgh (2011)) or the serial correlation and volatility of prices and quantities within and across metropolitan areas (e.g., Glaeser et. al. (2011)). Related work includes Arce and Lopez-Salido (2011), Burnside, Eichenbaum and Rebelo (2011), Lai and Van Order (2010), and Wheaton and Nechayev (2008).

[2] A large body of work also has been published about the consequences of the bust and government policy responses. Much of the scholarly analyses focuses on the subprime sector (e.g., Bajari, Chu and Park (2008), Danis & Pennington-Cross (2008), Demyanyk and Van Hemert (forthcoming), Gerardi, Shapiro and Willen (2007), Goetzmann, Peng and Yen (2009). Mayer and Pence (2008), Haughwout, et. al. (2011)), mortgage securitization (e.g., Bubb and Kaufman (2009), Keys et. al. (2010)), the default/foreclosure crisis (e.g., Adelino, Gerardi and Willen (2009), Campbell, Giglio and Pathak (forthcoming), Foote, Gerardi and Willen (2008), Gerardi et. al. (2008), Mayer, Pence and Sherlund (2009), Mian and Sufi (2009), Mian, Sufi and Trebbi (2010), Piskorski, Seru and Vig (2009)) or the role of government regulation (e.g., Avery and Brevoort (2010), Bhutta (2009), Ho & Pennington-Cross (2008)).

[3] For example, some of the work cited above uses newly available micro data on subprime loans, but this does not cover all home purchases or go far enough back in time to capture early booms (as we show below). There are aggregate price series at the national level and for some metropolitan areas that go well back in time, but no high quality neighborhood-level have been available heretofore.

In this paper, we provide the first estimates of the location, timing, magnitude, and potential causes of the beginning of the housing boom at the MSA and neighborhood levels in the United States. We use the price growth rate series in each MSA and neighborhood to look for housing booms. More precisely, the start of the boom is estimated as the global structural break in the relevant area's quality-adjusted house price appreciation rate series. Price growth rates are used instead of price levels due to important features of housing markets: (a) the existence of search frictions and credit constraints that are thought to prevent the immediate capitalization of changes in fundamentals into price levels (Wheaton (1990); Krainer (2001); Piazzesi and Schneider (2009)); (b) the empirical fact that house prices do not follow a random walk, as evidenced by the significant short-term persistence and longer-term mean reversion of prices (Case and Shiller (1989); Cutler, Poterba and Summers (1991); Glaeser, et. al. (2011)); and (c) potential irrational sentiment about price growth rates (Shiller (2005).

Our estimates rely on a new micro data set which contains over 23 million housing transactions in 94 metropolitan areas from 1993 to 2009, provided by DataQuick. It contains the exact date of the transaction and the precise location of each house, which is critical to the construction of neighborhood-level price indexes. The large micro data set also enables us to use a split sample approach (as in Card, Mas and Rothstein (2008)) to solve the specification search bias problem in the estimation of the magnitude of price breakpoints (Leamer (1983)). We further match this transactions data with individual loan information from the Home Mortgage Disclosure Act (HMDA), which allows us to observe household income of all accepted and rejected loan applicants (along with various demographic features), both at the metropolitan area and neighborhood levels. This allows us to investigate whether key fundamental drivers of housing demand also change around the time when prices start to boom.

Our analysis generates five important findings. First, there is wide heterogeneity in when the boom began in different areas. We find major jumps in price appreciation rates as early as 1997 in some MSAs and as late as 2005 in others, with only a modestly higher concentration of breakpoints around 2004-2005. Our neighborhood analysis – based on groups of 2 or 3 contiguous census tracts – also shows similar heterogeneity in timing, with booms beginning as early as 1996 and as late as 2006. Thus, Figure 1's plot of the two most prominent aggregate house price indexes provides a misleading sense that the recent boom was a single national event which began in the early to middle part of the last decade. These results also highlight how

3

problematic it can be to analyze data that has been pooled across markets at a given point in time. Doing so effectively averages across locales which are at very different points in their housing boom cycles.

Second, the magnitude of the initial jump in price growth at the start of the boom was economically, not just statistically, significant. After controlling for time and location fixed effects, log house prices are over 4 percentage points higher on average during the first year of the boom relative to the previous 12 month period both at the MSA and neighborhood levels.[4] In addition, there is no evidence that home price growth was trending up prior to the start of the boom. Moreover, our analysis shows that neighborhood data is critical to properly estimate the magnitude of the boom in some areas such as San Francisco, where there is wide dispersion in the timing of booms by neighborhood. However, this is not the case in all areas, as illustrated by Las Vegas, where more than three quarters of the neighborhoods boom in a single six month period.

Third, of all potential demand shifters examined, ranging from income and demographics to credit market variables, only local incomes have an economic and statistically significant jump around the time of the global breakpoint in house price appreciation. The magnitude of the income change can account for half or more of the initial jump in price appreciation on average, even without appealing to forward looking expectations about future income growth.[5] Like price growth, income does not exhibit any trend prior to the beginning of the boom and it continues growing after the boom has started, but at a much lower rate than housing prices. While these estimates suggest that the beginning of the boom was fundamentally justified on average, they do not imply that what followed was rational, as our analysis pertains strictly to the start.

Fourth, there is important heterogeneity in price and income changes across areas at the beginning of their booms. For example, price growth changes are higher than average among the one-third most elastically-supplied markets in our sample (usually late booming markets), but there is no jump in income when their housing booms begin, so there is no evidence that any fundamentals played a key role in those places. On the other hand, income growth is nearly as large as price growth around the global breakpoints for early booming and inelastically-supplied

---

[4] See equation (5) and the discussion in the text for more details on this result.
[5] See section V for details.

markets, so virtually all of their booms can be accounted for by local fundamentals (given reasonable assumptions about the income elasticity of demand for housing).

Fifth and finally, none of the other potential demand shifters have materially positive pre-trends either, but some do lag the start of the boom. For example, in the second year after the boom there is a reduction in the fraction of government-insured loans targeted towards low-income households (as reflected in the share insured by the Federal Housing Administration (FHA)), and an increase in the market share of private subprime lenders. Since lending to low income homeowners and higher risk borrowers do not exhibit pre-trends or positive jumps at the breakpoint, it seems that key players in the lending market responded to the initial jump in price appreciation rates, rather than caused it. This suggests that future work may address the question of whether the nature of subprime loans amplified the bust versus the possibility that many of these loans simply were issued close to the end of the booms in many markets.

Our empirical strategy is better designed to understand how factors with local or regional variation lead or lag the housing boom. This is relevant for the interpretation of the role of mortgage interest rates, which tend to be quite similar across the country. Another caveat is that we do not investigate how unobserved factors, such as contagion or the influence of irrational expectations, contribute to the housing boom. These factors may have been non-trivial, especially in places where we find no evidence that any demand shifter jumped around the time the boom began. Hopefully, future research will fruitfully use the timeline of the beginning of the housing boom first provided in this paper to investigate their potential roles in the start and subsequent development of the boom.

The rest of the paper proceeds as follows. The next section briefly outlines the conceptual framework within which house prices and house price growth rates can be analyzed. Section III follows with a description of our data sources. Section IV then describes how we create price indexes at the metropolitan area and census tract group/neighborhood levels. The process by which global breakpoints in the price appreciation rate series are estimated also is documented here. Results on the timing and magnitudes of the price growth jumps are then presented. Section V contains the analysis of potential causes of housing booms. Section VI concludes and presents a brief discussion of future research.

II.   A Conceptual Framework for Evaluating House Prices and the Beginning of a Boom

Theory does not provide a clear guide about how to define the beginning of a boom. Empirically, one could use a jump in the level of prices or in their growth rate to determine the timing of the starting point. A standard asset pricing approach suggests that changes in price levels are more appropriate, as any given fundamental shock typically results in a discrete jump to a new price level, with no impact on the asset's return.[6] However, we define the beginning of the boom as occurring when there is a global breakpoint in the relevant area's rate of price appreciation. This is done for a variety of reasons.

One is the nature of the data. House prices do not follow a random walk. Rather than sharp jumps in price levels followed either by flat prices or values growing at a steady rate over time, there are long cycles of varying amplitude as shown by the plot in Figure 1. Case and Shiller (1989) and Cutler, Poterba and Summers (1991) long ago documented the predictability of house prices in the form of economically and statistically significant short-term persistence and longer-term mean reversion, with Glaeser, et. al. (2011) providing a recent update.

Search models with bilateral negotiation between buyers and sellers provide one possible explanation for why changes in fundamentals need not be immediately reflected in price levels, but can result in changes in price growth over time (e.g., Wheaton (1990); Krainer (2001); Piazzesi and Schneider (2009)).[7] More generally, different amounts of price volatility can arise from search models depending upon their underlying vacancy conditions and the precise nature of the search process.[8] Lagged supply responses also may help account for the slow adjustment of prices over time, as well as provide rationales for different price growth depending upon what one assumes about a natural rate of vacancy (e.g., Rosen and Smith (1983)). The presence of credit constraints also has been shown to lead to shocks that are reinforcing, generating potentially high price volatility across hot and cold markets (e.g., Stein (1995); Krainer (2001)). There also could be non-rational factors that justify using changes in the rate of price growth to define the time line of a boom. In this regard, Shiller (2005) argues that some type of irrational

---

[6] Of course, the capital gain does not equal the total return on housing. However, we do not readily observe the implicit rental flow on owner-occupied housing. Holding constant the service flow, the basic argument still holds.
[7] In these models, which typically are short-run in nature in the sense that the stock of housing is fixed, people become mismatched from their current housing units for a variety of demographic (e.g., marriage, divorce, having children) and other reasons (unrelated to market fundamentals) that result in a discrete change in the demand for a given quality housing unit. As long as these households cannot immediately be matched with another appropriate housing unit, any change in fundamentals will not be fully and immediately reflected in market prices.
[8] See Wheaton's (1990) seminal article for some examples. Even without search, there are good reasons not to expect straightforward jumps from one equilibrium price level to another. Poterba's (1984) classic capital markets model of house prices shows that overshooting is to be expected, with a transition to steady state equilibrium.

exuberance simply allows individuals to form incredible expectations about the future rate of price growth.[9] Whatever the cause, there is good reason to focus on changes in price growth rates, not price levels, in determining the timing of the beginning of a boom.[10]

Once we have determined the timing and magnitudes of the beginning of the different local booms, we then investigate potential causes of those initial jumps in price growth. The starting point for that analysis is the core model of spatial equilibrium from urban economics (Rosen (1979); Roback (1982)). That model views house prices as the entry fee necessary to access the underlying productivity and amenities of a local labor market area. It presumes a reference locality that offers a utility level, U*, which defines the minimum level of utility that can be obtained anywhere. Mobility ensures that utility is equated across markets at this level, so that nobody has an incentive to relocate in equilibrium.

A highly simplified, linearized version of this framework would depict utility as the sum of wages (W, which reflect local productivity) and amenities (A) less the cost of housing (R for rent, which in the traditional model posits absentee landlords). That is,

(1) $U^* = W + A - R.$

With utility fixed at U*, simple differentiation yields

(2) $dU^* = dW + dA - dR = 0$ or $dR = dW + dA.$

This equation indicates that house prices (or rents) should increase in places with positive shocks to wages and amenities.[11]

This basic insight of the long run equilibrium serves as a useful starting point for our investigation of what factors are correlated with the beginning of booms (see Section V), although for that part of the analysis, we adopt the following perspective: in the short-run, any

---

[9] Piazzesi and Schneider (2009) introduce another potential non-rational factor in the form of momentum traders in their search model.

[10] Another reason we use the jump in the rate of growth, rather than the level, of prices is that one-time changes in the levels can be seasonal in nature.

[11] This model, which serves as the foundation for much empirical work on the spatial distribution of house prices, is empirically powerful as indicated by the $R^2=0.60$ resulting from a very simple regression of log house prices on log median family income and mean January temperature in a cross section of metropolitan areas (Glaeser and Gyourko (p. 348, 2005)).

7

demand shifter could help account for the beginning of a boom because supply is likely to be highly inelastic over such a narrow window of time. Thus, we expand our analysis beyond the typical income and amenity factors suggested by the Rosen/Roback model to include credit market variables and various demographics that are likely to be correlated with the demand for housing. Before that, however, we briefly describe the data used in the empirical work in Section III and then estimate the timing and magnitudes of the beginning of local booms in Section IV.

III.   Data Description

Our primary source of housing market data comes from DataQuick. This firm provides micro observations on home purchases collected from deed records. Data are available for at least one county in 269 MSAs, although the starting point for this information ranges from the late 1980s to very recently. We observe the exact date, price, and location (street address plus census tract identification) of each housing transaction, in addition to a variety of housing traits such as the number of bedrooms, number of bathrooms, square footage of living space in the home, and the year the unit was constructed. 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).

The final sample of DataQuick observations used in this analysis contains more than 23 million arms-length, single-family housing transactions in 94 metropolitan areas spread across 29 states, from 1993(Q1) to 2009(Q3). 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.[12] The online Appendix #1 provides a list of all MSAs in our sample, and their respective starting dates.[13]

Table 1 reports summary statistics based on the 2000 Census for the nation, the 269 markets tracked by DataQuick, and our final sample of 94 MSAs. Southern metropolitan areas

---

[12] We also conducted all the analysis reported below on a larger sample of 112 MSAs. The additional MSAs had imperfect data on housing characteristics or recorded transactions that only start between 1998 and 2000. Average estimates were very similar for both samples.

[13] All appendices are on-line and available from our individual websites.

are underrepresented and western markets overrepresented compared to the nation.[14] Typical house value is higher in our sample, largely because of the overrepresentation of west coast markets. The same is true for income, although the differences are not so large for this variable. Other demographics such as race and educational achievement show similar values across the different samples reported in Table 1.

Further detail on prices, quantities, and select housing traits at the metropolitan area and neighborhood levels (proxied by tract groups) is provided in the table in on-line Appendix 2. We pool contiguous tracts into tract groups to provide sufficient observations to estimate price indexes at the neighborhood level. These contiguous census tracts were combined into pairs (and sometimes triplets when necessary) using a random process to form tract groups. Average characteristics by MSA (weighted by the total number of transactions) are quite similar to averages by neighborhood. For example, the average price (weighted by population) by MSA is about $255,000, while the average price by neighborhoods is nearly $257,000. The main difference between neighborhoods and MSAs obviously is in the total number of transactions: MSAs, on average, have around 250,000 transactions in our data, while neighborhoods have just over 1,700. Because neighborhoods have a much smaller number of transactions, we restrict our neighborhood-level analysis to a sample of tract groups that have more than 10 observations per half-year.[15] That constraint mitigates the chances of estimating price indexes and breakpoints with substantial error. The average characteristics for the restricted sample are similar to those for the complete set of tract groups. Naturally, the average number of transactions is higher, as it increases to 2,260.

DataQuick provides no demographic or income-related information, so we merge their data with Home Mortgage Disclosure Act (HMDA) files in order to include information on the income of all loan applicants, in addition to race and various loan characteristics.[16] Various procedures were used to select a subsample of DataQuick transactions and HMDA loan

---

[14] A major reason for this is that Texas markets are not in the DataQuick samples. Western region metropolitan areas are overrepresented largely because of extensive coverage of California, where DataQuick started its business.
[15] We further restrict the sample of tract groups when estimating the magnitude of the break points using the split sample approach explained in section III.C. There, we impose the constraint that tract groups have more than 10 observations per half-year in both random samples.
[16] The Home Mortgage Disclosure Act (HMDA) was enacted in 1975, and implemented by the Federal Reserve Board. It requires that lending institutions report virtually all mortgage application and loan data. See http://www.ffiec.gov/hmda/ for details.

applications to ensure consistency and data quality.[17] Both data sets were then merged using a straightforward sequential matching process.[18] In total, 92.7% of all sales transactions in DataQuick were matched at some point in the sequential 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. Of the matched observations, 25% were randomly assigned due to having multiple matches.

IV. The Timing of the Housing Boom

IV.A. Price Index

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). For the final three variables, categorical dummies are created as described in our on line Appendix #3. The hedonic index values are derived from the coefficients in the vector α_6 on the year-quarter dummies ($YearQtr$) in the following equation:

$$(3)\ Log(HP_{m,t}) = \alpha_0 + \alpha_1*Bed_{m,t} + \alpha_2*Bath_{m,t} + \alpha_3*Age_{m,t} + \alpha_4*Sqft_{m,t} + \alpha_5*Sqft^2_{m,t} +$$
$$+ \alpha_6 YearQtr_t + \epsilon_{m,t},$$

---

[17] For the merge, we restrict each sample to transactions or loan applications pertaining to single family homes with a positive first lien amount. Because HMDA only reports property type and whether a loan was a first or subordinate lien starting in 2004, we cannot restrict to single family units prior to this year. [This information is reported for all years in DataQuick.] We separately merge accepted and rejected loans, as reported by HMDA. The rejected loan sample includes outright rejections, loan applications that were voluntarily withdrawn and those that were approved but not ultimately used to purchase a home. For the rejected loans, we assume that the pool of houses sold in DataQuick (and their associated lenders) was the relevant target for those non-approved loans. The rejected loan sample is used only to calculate local income, while the sample of actual sales transactions is used in the rest of the analysis. Finally, census tracts were all converted to 2000 boundary definitions for geographic consistency.

[18] 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. In the next step, unmatched observations where 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.

where $\epsilon_{m,t}$ is an idiosyncratic error term. The estimated indexes are then normalized to 100 in 2000(Q1) for all MSAs. When estimating a similar model for neighborhoods, we replace the year-quarter dummies by half-year dummies to increase precision, as each individual neighborhood naturally has a much smaller number of transactions in any given time period.

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).[19] We employ hedonic price indexes because their data requirements are much less onerous. This is particularly helpful when we create price indexes at the neighborhood level. To check the robustness of our estimates, Figure 2 plots the hedonic index from equation (3), the S&P/Case-Shiller repeat sales index, and a repeat sales index we created using the DataQuick files for the Las Vegas-Paradise metropolitan area.[20]

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.[21] Hence, our hedonic method captures the price movements tracked with other widely used methods to a very great extent.

Figure 3 then plots the individual hedonic price indexes over time for each of our 94 metropolitan areas. Prices tend to be relatively flat for a period of years prior to the great boom, and then show fairly long, multiyear swings. There is a noteworthy expansion in the variation of prices across markets during the boom, which has not disappeared as of the end of our data in 2009(3)). Figure 4 then displays the dispersion in price index levels across neighborhoods for four representative metropolitan areas. While the dispersion in prices across neighborhoods

---

[19] Their methodology is an update on the approach introduced by Bailey, Muth and Nourse (1963).

[20] 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.

[21] Of the 14 Case-Shiller markets that overlap with the DataQuick 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.

increased at least modestly in each area depicted, there is variation in its extent across metropolitan areas. For example, there is very little change in the degree of price dispersion across tracts in the Detroit area. As we discuss below, this is one of the markets that did not actually experience a significant jump in house price appreciation rates at any time over the past decade or two. In contrast, the increase in dispersion is quite large in the San Francisco area. Las Vegas and Orlando show increases in dispersion across neighborhoods, but the change is less stark in these markets.

IV.B. The Beginning of Local Booms: Breakpoint Estimation

For both MSAs and tract groups, the price growth rate series is constructed by dividing the estimated price index from equation (3) in period t by the estimated index in period t-4. This is done to address the large seasonality in home transactions. We then look for jumps in the resulting appreciation rate series. Because there can be multiple such jumps, we find the quarter during which there is a global structural break in the quality-adjusted price appreciation rate series.[22]

More specifically, we estimate the following equation for all potential structural breakpoints ($q^*_{m,t}$) for each MSA m and time t :

$$(4) \quad PG_{m,t} = a_m + d_m\mathbf{1}[q_{m,t} \geq q^*_{m,t}] + e_{m,t} \text{ for } T_{m,0} < q_{m,t} < T_{m,N}$$

where $PG_{m,t}$ is annual house price growth computed from the hedonic price indexes created as described above, d_m estimates the importance of the potential point break, $q_{m,t}$ is a quarter, $q^*_{m,t}$ is the location of the potential structural break, $T_{m,0}$ is the first quarter of data, and $T_{m,N}$ is the

---

[22] We also considered scaling growth rates of prices by some metric. One obvious candidate is rents, but data quality limitations prevent us from doing so. Owner-occupied and rental stocks are not very similar or well integrated in most U.S. markets, so it is difficult to credibly compare house prices and rents on the same properties (Glaeser and Gyourko (2010)). Moreover, rent data from REIS are available for only 29 of our MSAs, and are derived from surveys of private landlords and, thus, do not reflect observations on actual transactions. Nevertheless, we still use the REIS rent data as a robustness check to see if they exhibit a similar time pattern to house prices. We also examined construction costs. The available data reveals that these costs are quite stable compared to overall house prices (Gyourko & Saiz (2006)). Thus, the vast majority of the change in the ratio of price-to-construction cost is driven by house prices. Finally, we could use a real house price series, deflating our nominal data with national or regional CPIs. We do not do so explicitly because the quarter fixed effects included in the estimation already account for national changes in cost of living. Moreover, inflation generally does not change much over the short time periods compared in the analysis below (i.e., around the quarter of the breakpoint in the house price growth series).

quarter of the highest appreciation rate.[23] The estimated global breakpoint ($q^{**}_{m,t}$) maximizes the R^2 of equation (4). Essentially, the breakpoint occurs in the quarter in which the change in the price growth series has its greatest impact in explaining the price growth series itself.[24]

Summary statistics on the breakpoint estimations at the MSA and neighborhood levels are presented in on line Appendix #4. Overall, the average R^2 of the global break point is 62 percent for all MSAs and 49 percent at the neighborhood level. The average number of quarters from the start of the area's time series to the point of maximum price growth is just under 30 quarters, so we typically estimate the breakpoint using 7.5 years of data. Those results include the impact of 15 metropolitan areas that we conclude never experienced a boom in house prices.[25] For six metropolitan areas, we did not find a breakpoint estimate (d_m) that was statistically significantly different from zero. For the other nine of these 15 areas, there were less than twelve quarters of data in the estimation sample, so we could not reasonably estimate a breakpoint for these places. These tend to be markets with flat or negative price growth trends over the long run, so the quarter with the highest price appreciation often was close to the starting date of the sample. Among the 7,335 neighborhoods for which we have price data, 470 do not have at least twelve quarters of price data with which to estimate a global breakpoint, and another 1,728 are found to have estimated breakpoints that cannot be distinguished from zero at standard confidence levels.

For the remaining 79 metropolitan areas and 5,137 neighborhoods for which we did estimate statistically significant breakpoints in price growth, Figures 5 and 6 plot the distribution of starting points over time. The striking feature of these plots is how uniform their distributions are. Clearly, it is a mistake to consider what happened a single national event, much less one characterized by booming exclusively after 2002. While there is a concentration of metropolitan areas markets with booms beginning in the 2002-2004 period (43 metropolitan areas experienced their global breakpoints during this time frame according to Figure 5), 21 metropolitan areas saw

---

[23] We restrict the search for a structural break point within this time frame to minimize any influence of the bust period in the estimation of the beginning of the boom.

[24] See Hansen (2000) for a proof of the consistency of this procedure.

[25] All 15 are located in the interior of the country, with nearly half readily characterized as 'rust belt' markets. These seven include Akron, OH, Cincinnati-Middletown, OH-KY-IN, Cleveland, OH, Columbus, OH, Detroit-Livonia, MI, Erie, PA, and Springfield, OH. Another four are small to modest-size metropolitan areas in border or southern states (Knoxville, TN, Nashville-Davidson-Murfreesboro, TN, Oklahoma City, OK, Tulsa, OK). The remaining four include three Colorado markets (Colorado Springs, Denver-Aurora, and Ft. Collins-Loveland) and Lincoln, NE.

the beginning of their housing booms before the year 2000. And, another seven markets had their global breakpoints in 2000 or 2001, before the beginning of the big wave of booms. Finally, there were eight more markets there were very late boomers in 2005. The distribution for neighborhoods in Figure 6 is even more uniform than that for metropolitan areas.

IV.B.i. Geographic heterogeneity in the beginning of the boom

The geography of the timing of the start of housing booms is interesting in its own right, and is suggestive of a role for contagion in explaining the spread of the boom across markets. Because this paper focuses exclusively on the beginning of the boom, we report data on geographic heterogeneity at the metropolitan area level in on-line Appendix #5. Very briefly, the first booms at the metropolitan area level occurred in northern New England (Massachusetts and Connecticut) and coastal California. On the east coast, they then spread east and south from northern New England. They spread north and east from coastal California.

Figure 7 plots the distribution of tract group-level breakpoints over time in four individual metropolitan areas. Some major coastal metropolitan areas, such as San Francisco, show a wide range of breakpoint times across neighborhoods that are almost uniform in nature. This is in stark contrast to what happened in Las Vegas, where at least three-quarters of their neighborhoods boom in one six month period (the first half of 2004). The Orlando market is somewhere in between these extremes, with almost 40% of its tract groups booming in the same six-month period. A similar middle of the road pattern is seen in Detroit, with the caveat that, by our definition, only 11 percent of its neighborhoods had a boom. This heterogeneity will be relevant to the issue of how differences in the timing of booms at the neighborhood level can impact measurement of the magnitude of booms at the metropolitan area level (which is discussed more fully in the next section).[26]

IV.C. How Big Were the Booms? Magnitudes of the Structural Breaks

That there were statistically significant jumps in appreciation rates in most markets at some point in time begs the question of whether these structural breaks were economically

---

[26] The analogous point applies with respect to the variation across metropolitan areas and measurement at the national level.

significant, too. Our basic estimation strategy for determining magnitudes is based on the following equation:

$$(5) \quad PG_{m,t} = \psi(q^{**}_{m,t}) + \kappa_t + \alpha_m + e_{m,t,\tau} \, ,$$

where $PG_{m,t}$ is the annual housing price growth in each MSA, $\psi(q^{**}_{m,t})$ is a set of relative year dummies (with the dummy for the twelve month period prior to the estimated structural break—$q^{**}_{m,t}$--excluded from the estimation; thus, the year of the breakpoint includes the quarter of the structural break and the three subsequent quarters), κ_t are year-quarter fixed effects, and α_m are MSA fixed effects.[27] We chose a 12-month horizon to gauge magnitudes of the jumps because there is noise in the quarterly data which is not solely related to estimation error. For example, when home prices are agreed upon and when they close can vary by a quarter or more, depending upon how long it takes to arrange financing and solve any possible outstanding legal issues. In addition, prices in housing markets move slowly and are positively serially correlated over short horizons as discussed above.[28]

Whenever estimating the magnitude of the jump at the breakpoint, we follow Card, Mas and Rothstein (2008) in employing randomly drawn split samples of houses, with one sample of houses used to estimate the time of the breakpoint and the other used to estimate the magnitude. Such an approach mitigates potential specification search bias as described in Leamer (1983).[29]

Our estimates are summarized in Table 2. The first column reports the magnitude of the jump when no time or location fixed effects are controlled for. Price growth is about 7.5% higher at the breakpoint at both the MSA and neighborhood-levels in that specification. Column 2's numbers are from a specification including time dummies, so they represent the variation

---

[27] The inclusion of time fixed effects in equation (5), but not (4), is due to our interest in knowing how much of the variation in prices can be accounted for by local fundamentals. See Section V for more on this.

[28] That said, we also estimated magnitudes of the jump (with split samples) using a regression discontinuity approach that relies on the precise quarter of the break. That estimating equation is given by

(5a) $PG_{m,t,\tau} = d_m \mathbf{1}[q_{m,t} \geq q^{**}_{m,t}] + P(\tau_{m,\tau}, \gamma) + \kappa_t + \alpha_m + e_{m,t,\tau} \, ,$

where $PG_{m,t,\tau}$ now is the annual housing price growth in each MSA in the quarter relative to the estimated break point ($\tau = q_{m,t} - q^{**}_{m,t}$), d_m is the magnitude of the break point, $q^{**}_{m,t}$ is the estimated structural break, $P(\tau, \gamma)$ is a polynomial in τ (quartic, allowing for different slopes before and after estimated break point), κ_t are year-quarter fixed effects, and α_m are MSA fixed effects. Estimates of the magnitude of the jump in price appreciation rates are slightly smaller than those reported below in Table 2 (by about 20%), but each remains highly statistically significant.

[29] The estimated timing of the breakpoints in the full sample has a correlation of 0.82 with the analogous estimates that use a random split sample. That correlation increases to 0.91 once we constrain the sample to MSAs that had a statistically significant break point using the full sample.

remaining after sweeping out the impacts from nationwide changes that are common across all locations. The estimated jumps are lower, ranging from 5.4% at the MSA level to 6.2% at the neighborhood level. Column 3's results are from a specification that also includes area (MSA or neighborhood) fixed effects. Note that there is little change from column 2, indicating that there is plentiful local variation remaining. Henceforth, we analyze the variation in the magnitudes of price growth jumps which remains after controlling for national and local area fixed effects. The remaining two columns of Table 2 report results that either drop the quarter of the breakpoint or areas without statistically significant breakpoints. Because we are most interested in documenting and evaluating those places with statistically significant breaks, we focus on the results from column 5.

Figure 8 plots the MSA level results from column 5 of Table 2, including data from two years before and three years after the boom begins. The dashed lines mark the 95% confidence interval. Because everything is measured relative to the pre-breakpoint year, price growth for that year is zero, by definition. Note that there is no pre-trend in price appreciation. This is not a statistical artifact of the estimation procedure, so it appears that the jump in prices is an event that just happens at a given point in time, rather than building over time. The higher level of appreciation is maintained for another year before starting to decline. On average, it takes five years after the break year to get back to the level of price appreciation that existed before the break.

The plots in Figure 9 provide some insight into the extent of heterogeneity across metropolitan areas in the magnitudes of the price growth jumps and the time patterns of appreciation relative to the breakpoints. The first plot for the Detroit metropolitan area is included to illustrate the pattern typical of markets that did not have a significant structural breakpoint. The next two plots for the Las Vegas and Orlando metropolitan areas depict larger jumps (20 and 10 percentage points respectively) at the estimated breakpoint. Note that price appreciation falls precipitously in Las Vegas in the 2nd year after the breakpoint, while in Orlando it accelerates by another ten percentage points in the subsequent year before plateauing and then rapidly declining. The remaining plot shows the San Francisco metropolitan area had a more modest jump, but that could be downward biased because neighborhoods in that MSA had a more uniform set of starting points over time (see Figure 7). Even if each neighborhood in San

Francisco had large price growth jumps, the pooled data at the MSA level could show a flat trend in price growth rates.

The second row of Table 2 reports average estimates that are the neighborhood-level analogues to the metro-level results, with Figure 10 plotting the findings. Neighborhood-based point estimates are slightly larger, take longer to converge to pre-trend levels, and have overall increases in precision. Figure 11 then reports neighborhood-level results by MSA. Estimates are similar to MSA level results for Detroit, Las Vegas, and Orlando. The main difference is for San Francisco. As suggested just above, the neighborhood-level data for this metropolitan area show that it has a bigger breakpoint jump, and then a flat trend over the next five years.

V.   Potential Explanations of the Boom

We next investigate how a number of potential demand shifters are correlated with the timing and magnitudes of the beginning of local housing booms. The variables investigated include traditional income measure suggested by the Rosen/Roback model of spatial equilibrium[30], along with other variables that have been suggested by some as potential causes of the boom. Most prominent in this group are credit market factors such as subprime mortgage activity and federal low income-housing mortgage programs (FHA). One omission from this latter group is interest rates, as our empirical strategy is best suited to examining factors that vary across space and time. Mortgage rates are roughly constant across local markets and are best examined by other approaches.[31]

Table 3 first documents what was happening to log prices. Log prices not only provide a robustness test to our magnitude results, but they are also more convenient for two reasons: they reduce the importance of outliers, and provide an interesting descriptive of variation in prices for many years after the boom. Metropolitan averages are provided in the first column. The next columns under the heading "pre-Trend" report how log home values at the MSA and neighborhood levels were changing relative to the twelve month period leading up to the quarter in which the boom starts. That annual period just prior to the global breakpoint is the baseline year relative to which everything is measured (it is called time *0* in the plots below). The -0.014

---

[30] Because most local amenities do not change at all (e.g., weather) or only slowly over time (e.g., school quality), we focus on the potential role of income from this model.

[31] For examples, see recent work by Favilukis, et. al. (2011) and Glaeser, et. al. (2011), which reach somewhat different conclusions about the likely importance of the role played by rates in accounting for the development of the most recent boom. Neither focuses on the start of the boom.

value listed for MSAs implies that prices fell slightly, by 1.4%, relative to our baseline twelve month period. However, this change is not significantly different from zero, so there is no economically or statistically significant pre-trend in price growth on average prior to the boom. Neighborhood level estimates are statistically significant, but the point estimate has the opposite sign of what would be a potentially worrisome (for our purposes) pre-trend. Thus, there is no evidence that prices were trending up prior to our estimated breakpoints.

Columns (4) and (5), labeled 'breakpoint jump', then report what was happening to log prices relative to our baseline year in the twelve months beginning with the quarter of the start of the boom (time *+1* in the plots below). For MSAs, prices were 4.1% higher, while the point estimates for neighborhoods is 4.4%, with both being statistically different from zero.[32] The final columns, labeled "2 years after break', report log prices during the next relative year (this is called time *+2* in our plots). The data here show that house prices clearly kept growing after the boom began, as indicated by the fact that they were 9.3% higher relative to the baseline year for both MSAs and neighborhoods.

V.A. Incomes

The next row of Table 3 documents whether income – measured by the log income of all local loan applicants - shows similar patterns. The strength of this variable is that it reflects the household income that was reported on mortgage applications to buy local housing, so it well proxies the income of potential demanders of local housing. Importantly, it also has quarterly variation, so it can be precisely calculated around the breakpoint. In addition, estimates of income changes at the breakpoint denoting the beginning of the boom seem less likely to suffer from reverse causality, especially given that we know there is no positive pre-trend in prices. This type of endogeneity could be a more significant concern if we were using income only of homebuyers, so it is important to include the income of all potential demanders for local housing. About one-third of the loans in HMDA are rejected, withdrawn, or approved but not accepted, so this measure is based on a pool considerably larger than just the actual homebuyers.[33]

---

[32] Note that these estimates are slightly smaller than the 6% that we estimated using actual price growth rates as reported in Table 2. The logging of prices appears to be diminishing the impact of outliers, as is common with housing data. In fact, when trimming the top and bottom 1% of the price growth distribution, we find point estimates of around 5% for Table 2.

[33] Nonetheless, this variable still is not perfect, although we know of no sure way to improve upon it. For example, it does not include data for households that truly are in the pool of demanders, but never applied to get a loan.

Income averaged $76,881 in our sample. The 'pre-trend' column shows that homebuyer income was flat when compared to the baseline year, so there is no pre-trend in either prices or income. Local income jumped 2.5%-3.3% around the time the boom began, so that both prices and incomes jump significantly and at somewhat similar rates in the year of the breakpoint. The last column of Table 3 shows that incomes were 4.3% above their baseline levels two years after the break using the MSA data (4.4% at the neighborhood level), so incomes also kept rising after the boom started, but not as much as house prices. While we are most interested in the comparison at the breakpoint denoting the start of the boom, Figure 12 summarizes the results in Table 3 by plotting the price and income point estimates from two years prior to the baseline year to five years after it, for both MSAs and neighborhoods.

While house prices and incomes at both the metropolitan area and neighborhood levels are increasing simultaneously at the start of the boom, we have to make further assumptions to provide insight into whether the change in income might be large enough to account for most of the boom in prices. This can be done using results from the empirical literature on the income elasticity of demand for housing in combination with an assumption that the supply of housing is fixed in the short run. Allowing demand to shift out from growth in income in the face of no new supply produces an upper bound impact on prices. The housing economics literature suggests that the income elasticity of demand ranges from 0.75 to 1.[34] Applying these parameter values to the reported changes in log incomes during the breakpoint year indicates that a 3.3%

Included in this group are homebuyers that made all cash purchases. The income – and changes in income - of the households that are not in our sample can be lower, equal, or higher than the income of loan applicants, so it is difficult to sign a bias in our measure. One strategy is to use an even broader measure such as per capita income. We believe this is inadvisable because it clearly reflects income conditions among many households who are not potential demanders of owned units. Moreover, this variable suffers from two very serious drawback given our empirical strategy—namely, it is only available annually and at the MSA level. The former biases down its own volatility and its temporal correlation with house price growth. The latter means we cannot use it at the neighborhood level, where we have rich price data. Still, we do report a robustness test based on log median income in the next section.

[34] See Polinsky and Ellwood (1979) for a classic early examination that puts the elasticity in the 0.8-0.9 range. Subsequent research that generates 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)). We implicitly assume that the income changes we observe reflect movements in permanent income. To the extent they are not, one would not expect a positive correlation with house price changes in the short run. As a very simple robustness check on our own data, we ran a regression of log prices on log income interacted with relative year dummies, with time and MSA fixed effects included. The elasticity estimated for the 12 month period following the breakpoint was 0.75.

jump in local incomes could account for from 55% to 74% of the 4.4% jump in neighborhood house prices. The analogous range is 46% to 61% for metropolitan areas.[35,36]

That the beginning of the past boom is so strongly correlated with a factor such as local income that everyone would agree is a key fundamental of housing demand suggests a rational underpinning for the start of the boom on average. And, other tests indicate that conclusion is robust to outliers, to including other covariates, and to potential reverse causality.[37] However, these average results mask important heterogeneity in the relationship between house prices and incomes at the start of booms in different types of markets. It is to that issue that we next turn.

V.B. Heterogeneity in Prices and Incomes

Here, we systematically investigate the role of heterogeneity across markets along two dimensions: (1) timing of the boom (e.g., early versus late); and (2) elasticity of supply (e.g., highly inelastic versus highly elastic).[38] Table 4 reports our findings. Panel (a) shows how MSA and neighborhood prices and incomes changed before, during, and after the start of the boom for markets that were among the earliest and latest one-thirds of our sample to experience global breakpoints in price growth.[39] There was no positive pre-trend in prices in any of these groups. However, prices jumped during the breakpoint year and the increase in the late boomers was

---

[35] Dynamic characterizations of the Rosen/Roback framework indicate that income variability is sufficient to account for the growing heterogeneity in prices across housing markets over long periods of time (Van Nieuwerburgh and Weill (2010)), but that answers a different question from the one posed here.

[36] These calculations obviously ignore the role of unobservables such as potential changes in expectations. For example, Figure 12 for neighborhoods reports that incomes were 5.5% higher in the 3rd year after the breakpoint, and then become relatively stable after that. A forward-looking model that only captures the impact of the income growth in those years would imply an even larger influence for income on the beginning of the boom. However, extending this logic even further to include some type of expectations about the current bust period would work in the opposite direction. A complete analysis of the role of all potential expectations – and how rational or irrational they were - is well beyond the scope of this paper.

[37] On-line Appendix #6 reports these robustness tests in more detail. First, we show MSA-level figures of price changes against income changes at both pre-trend and breakpoint periods. The relationship between both variables is only strong at the breakpoint, and is not driven by outliers. A simple OLS regression of the price change variable on the change in income yields a coefficient of 0.68 (s.e. 0.09) and a R^2= of 0.37. Including covariates such as other likely demand shifters (e.g., percent minority, migration inflow, percent speculators, average loan-to-value ratio, percent subprime loans, and percent FHA loans), reduces the point estimate only modestly to 0.63(0.09), while the R^2 increases modestly to 0.48. Thus, there is no empirical evidence that other readily observable demand shifter can dramatically reduce the relationship between price and income changes. In addition, instrumenting homebuyer income with per capita income to deal with potential reverse causality does not change the results significantly.

[38] We also investigated two other sources of the heterogeneity: a) size of the MSA; and b) net demand (small increases in transactions relative to new supply versus large increases in that difference). However, we found no major differences on either dimension.

[39] The earliest one-third of booms started no later than 2000Q4. The latest one-third of booms started no earlier than 2003Q4.

greater than that of the early boomers (3.8% versus 3.0% for MSAs, and 6.6% versus 2.1% for neighborhoods). Prices continued to rise after the start of the boom, but more so for late boomers: two years after the boom began, prices were 5%-7% above those in the baseline year among the early boomers, versus 9%-14% above the baseline year in the late boomers.

Turning to income, we find no material evidence of a strong pre-trend in income growth by timing of the boom. However, the income jump relative to price jump was much greater among early boomers at the breakpoint. Among this group, the magnitudes of the income and price changes are almost identical, and even the lower end of the income elasticity of demand estimates would allow their income growth to account for virtually all of their price growth at the breakpoint. In stark contrast, the late booming markets experienced much smaller relative growth in income at the start of their booms. The change in income is about half that of prices, and it is not statistically different from zero for MSAs. We see much the same pattern by timing in the second year after the break. The results are even more striking at the neighborhood level: income tracks price movements in early booms during the first three years after the start of a housing boom; the price-income gap increases only in the fourth and fifth years.

Panel (b) of Table 4 documents that heterogeneity is not limited to early versus late boomers. It also differs by local supply conditions.[40] Once again, we see no material pre-trend for inelastically or elastically supplied markets. However, we find a much bigger jump in prices at the breakpoint among the most inelastically supplied markets: 3.8% versus a statistically insignificant 2.3% for the most elastically supplied markets. Among the one-third most inelastically supplied markets, the income jumps at the breakpoint are 65% of the magnitudes of the price jumps.

In sum, the material change in income we observe on average around the beginning of housing booms is not consistent across all types of markets. Income looks to be able to account for much of the start of the booms in early booming and inelastically supplied markets, but for little or none of the initial jump in price growth in late booming and elastically supplied markets.

V.C. Other Potential Demand Shifters

---

[40] We only have supply elasticity data for 76 MSAs. Using Saiz's (2010) measure, the one-third most inelastically supplied markets have elasticities below 1.1, while the one-third most elastically supplied markets have values greater than 1.7.

We also investigated whether a host of other factors that could affect the demand for housing also changed (and by how much) around the time the different local housing booms started. These include a variety of demographic and credit market variables, the results for which are reported in the top two panels of Table 5.

V.C.i. Demographics: The Pool of Buyers

Changes in the pool of buyers might be considered an amenity (or disamenity), as well as reflect a direct change in the demand for housing. Consequently, we analyzed how the share of minority and white purchasers changed around the start of the local housing booms, along with how the share of speculators changed. The "% minority" variable reflects the fraction of African-Americans and Hispanics as coded in the HMDA files. The "% white" variable is obtained from the same source. Our measure of speculators is comprised of all those who bought more than one home in the same metropolitan area between 1993 and 2009 *and* all those who sold a house no more than one year after the date of purchase.[41]

While there are statistically significant changes in many cases reported in this panel, none is economically important as a comparison of the changes with the variable means (reported in the first column) show. Thus, none of these variables appears to be related to the timing or magnitude of the beginning of the boom in house prices in an economically important way. In fact, the share of minority homebuyers was decreasing slightly prior to and at the breakpoint, while the fraction White was increasing. Given the high correlation between race and income in the U.S., this is not particularly surprising after the income results shown above. The same holds for our measures of speculators. There are only small and imprecise increases in the fraction of speculators, before, during, and after the breakpoint.

V.C.ii. Credit Markets

The next panel shows results for a number of credit market variables: the fraction of subprime loans, the fraction of FHA-insured loans, the average loan-to-value ratio (LTV) among homebuyers, and the mortgage interest rate. The decline of underwriting standards has been widely discussed as a key factor in the boom, and perhaps no other facet of those standards has been studied more than subprime mortgages. There is no one formal or legal definition of what

---

[41] See Bayer, Geissler and Roberts (2011) for another measure of speculation.

constitutes a subprime loan, and many papers use FICO scores to define what constitutes a subprime loan. Because we do not have credit bureau information, we cannot measure subprime shares in that way. Hence, we use information on the underlying lenders from the DataQuick files. 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.[42] Our subprime measure is the share of mortgages issued by these top twenty lenders.

Note that no jump in subprime shares is observed the year before or during the twelve month period after the price appreciation breakpoint. Hence, it is hard to imagine how this variable could possibly help account for the start of the different local housing booms. However, subprime share does increase two years after the breakpoint, although not by a huge amount (it is 0.9% higher than in the baseline year, on a mean of 15%). It does keep increasing beyond this time (see the figure in on-line Appendix #6), achieving a total increase that is 20% of the baseline share prior to the breakpoint. Thus, the rise in subprime share lags the beginning of booms by a significant amount of time, suggesting that this factor could have played a meaningful role in the future development of the extent of the boom, but not in initiating it.[43]

The fraction of FHA loans (which are separately identified in DataQuick) describes the time pattern of a governmental insurance program that targets low income borrowers. It shows small declines before, during, and after the breakpoint. FHA loans and subprime lenders constitute around a quarter of the lending market. The remainder consists of traditional conforming loans that generally are guaranteed by Freddie Mac and Fannie Mae and private jumbo lenders. Given the results for subprime and FHA shares, it is unlikely that both government and jumbo loans had major changes around the breakpoint.[44]

Average loan-to-value (LTV) ratio, also based on DataQuick, is another potentially important underwriting standard that could have changed in a manner that might explain the initial boom in house price appreciation rates. This measure reflects the total amount of mortgage debt on up to three loans recorded by DataQuick. LTV shows small, insignificant, and

---

[42] Previously, this was called *B&C Mortgage Finance*. See Chomsisengphet & Pennington-Cross (2006) for more detail on these lenders and lists.

[43] Using a different modeling strategy, Brueckner, Calem, and Nakamura (2011) argue that subprime lending is a consequence rather than a cause of bubble conditions.

[44] It is conceivable that jumbo loans increased market shares a few years after the break, given the gigantic increase in home values, but that is an issue for future research.

negative point estimates before and during the start of the boom. The slight increase in LTVs two years after the boom began is miniscule in economic terms. Thus, this credit market factor does not lead the boom, does not jump coincidently with the start of the boom, and unlike the subprime share, does not increase much after the boom begins.

Finally, the last credit market variable is the quarterly national 30-year fixed-rate mortgage interest rate collected from the Freddie Mac Primary Mortgage Market Survey. It does not vary across MSAs, so we modify equation (5) to drop quarter and MSA fixed effects. Overall, mortgage rates fell by about 34 basis points in the year prior to the baseline year. Rates were lower in the breakpoint year than in the baseline year, but they were higher than before the baseline (by about 20 basis points, or 34-14). They fell slightly by about another four basis points on average by the second year after the breakpoint. There is nothing in this pattern to indicate that rates were likely to have caused the boom, as no reasonable elasticity of housing demand with respect to interest rates would generate anything approaching the 4.1 point rise in prices at the breakpoint.[45] However, the lack of variation across locations for this variable makes it hard to interpret in our context.

V.D. Robustness Checks: Other Price and Income Measures

The results for annual rents are based on surveys from REIS and are only available for a small subset of 29 MSAs. Tthere is no evidence of any pre-trend in rents. Rents are 1.4% higher in the breakpoint year relative to the baseline, but this average is not statistically different from zero, as indicated by the standard deviation of 1.1%. Rents do continue growing after the start of the housing boom, as indicated by the fact that they are 2.5% above baseline levels in the second year after the boom (final column). Given the survey-based nature of this series, relatively little volatility is to be expected. We interpret the fact that the time pattern is quite similar to that of house prices as supporting the internal validity of our findings on house prices, since it would be unlikely to have a housing boom without spillovers to the rental market, even given the caveats

---

[45] As noted above, there is a debate in the literature on the magnitude of this elasticity (see Himmelberg, Mayer and Sinai (2005) and Glaeser, Gottlieb and Gyourko (2011), respectively, for upper and lower bound estimates), but given the small change in rates at the breakpoint, not even the figures from the high end of the range would lead one to believe that lower rates caused much of the start of the boom. There was a much larger decline in rates in the five years preceding the start of the boom – see Appendix #6 - but given the lack of a pre-trend in house price growth, that too provides no evidence for a strong interest rate effect on prices leading up to or at the start of the boom.

noted earlier about these markets not being very well integrated in many regions of the country and the survey-based nature of this series.

As suggested above, the two alternative measures of household income reported on in this table are likely to generate biased results. The estimated impact of homebuyer income on price changes is upward biased if there is reverse causality, i.e., if incomes are higher only because high income families are the only ones that can afford the more expensive homes. Per capita income, on the other side, is downward biased if it includes a large fraction of families with stable incomes that do not demand owned homes (such as students, low income families, etc.). The downward bias in per capita income is aggravated by the lack of neighborhood and quarterly variation in that data. Our estimates confirm both intuitions. Each alternative income measure shows large jumps at the breakpoint and two years later, but the homebuyer income jump is higher and the per capita income jump is lower than our preferred measure of local income.

V.E. Quantities: Transactions Volume and Net Migration

Table 5 also reports on two sets of transactions measures. One is the total number of home sales, as recorded in the DataQuick files. The other is the annual net inflow of population based on IRS county migration data.[46] Housing booms appear to be preceded by increases in demand for MSAs and neighborhoods, as both of these variables show a positive pre-trend, although the estimates for net inflows of population are economically small.[47] Both variables are lower in the breakpoint year, although they still remain above baseline year levels. Two years after the boom, the total number of transactions has fallen further to only 1.2% more than in the baseline year, and net migration has fallen below that in the baseline year prior to the start of the boom. In sum, these results show a general pattern of trading volumes increasing prior to the start of the boom, but there is no evidence of discrete jumps in the year the price boom begins. And, these volume measures continue to fall, not rise, in the years immediately following the beginning of the boom.

---

[46] We thank Todd Sinai for providing these data in easily usable form.
[47] For example, an MSA of 300,000 people would have seen a net inflow of only 669 people in a year. This results from multiplying the migration coefficient of 0.223 by 300,000 and dividing by 100 (the units of the migration variable).

V.F. New Supply and Construction Costs

Table 5 also reports two measures of new supply. One is for new homes, which are counted in the DataQuick files. The other is for total building permits as reported by the Bureau of Commerce. This latter variable is only available at the metropolitan area level. Both new supply measures are higher two years before the boom than they are in the baseline year, although neither is statistically significant. A slightly greater number of new homes are being supplied in the breakpoint year relative to the pre-trend year (3.0 versus 2.4 per 100 people), so there is no large economically or statistically significant change in the breakpoint year. New supply is slightly higher two years after the start of the boom (at 3.6 homes per 100 people), but once again there is no evidence of a huge change. The permits data show much less variation over this time period.

Next, annual construction costs, which are estimates for a typical new home in each market based on data from the R. S. Means Company, are quite stable in the years around the breakpoint. As with the rent data, this series is based on underlying surveys and tends to be quite smooth over annual horizons.

V.G. Housing Traits as Placebos

Finally, the last two rows of Table 5 report the changes in the square footage and number of bathrooms in the homes in our sample before, during, and after the start of the boom. Homes were becoming very slightly larger before the boom, but the changes are small and not statistically significant. There is no jump in quality at the price breakpoint or afterwards.

VI.   Conclusions & Implications for Future Research

The national aggregate price data indicate that the dominant feature of the recent housing cycle was a truly great boom from 2003-2005. This very short period stands out even more in Shiller's (2005) longer-run aggregate price series. However the national data mask substantial heterogeneity across and within markets. Using the global breakpoint in the relevant house price appreciation rate series, we show that housing booms began in different markets (both at the metropolitan area and neighborhood level) as early as 1996/7 and as late as 2005/6. Thus, the recent boom was not a homogenous event that started everywhere at roughly the same time.

There also is much heterogeneity in the magnitudes of the initial jumps in price growth. On average, the MSA and neighborhood data show that the rate of price appreciation increased by about 4 percentage points, or more than 50% from the baseline nominal rate of growth just before the global breakpoint. However, early booming markets tended to jump by lower amounts, versus late booming markets which jumped up to 25 percentage points. Precise estimates of the price jump magnitude by MSAs are only achieved with neighborhood data, as some MSAs show great dispersion in breakpoints across their neighborhoods.

When we look at various potential causes of the boom, only income increases materially at the same time that the rate of price appreciation initially jumps. However, there is important heterogeneity in this effect, and a back-of-the-envelope calculation suggests that this fundamental can account for the majority of the magnitude in price jumps in early booming and in inelastically supplied markets, but for little or none of the jump in late booming and elastically supplied markets. A host of other factors, including much discussed underwriting standards as reflected in LTVs and subprime mortgage share do not increase prior to the start of the boom or coincidentally with the breakpoint in price appreciation. There is a substantial increase in subprime share after the boom starts, which suggests that this credit market factor more responded to the price increases than caused them.

Finally, these results establish the foundation for a host of future work exploring how the boom developed over time. The geography of the timing of the beginning of booms across markets raises the possibility of spillover effects and contagion based on proximity. Fully developing the time line of the cycle should help us better investigate the role of various factors in how the boom evolved and the bust ensued.

Select References

Adelino, Manuel, Kristopher Gerardi and Paul Willen. "Why Don't Lenders Renegotiate More Home Mortgages? Redefaults, Self-Cures and Securitization", FRB Atlanta Working Paper 2009-17.

Arce, Oscar and David Lopez-Salido. "Housing Bubbles", *American Economic Journal: Macroeconomics*, Vol. 3 (January 2011): 212-241.

Avery, Robert and Kenneth Brevoort. "The Subprime Crisis: How Much Did Lender Regulation Matter". Division of Research and Statistics. Board of Governors of the Federal Reserve System. August 2010.

Bailey, Martin, Richard Muth and Hugh Nourse. "A Regression Method for Real Estate Price Index Construction", *Journal of the American Statistical Association*, Vol. 58 (1963): 933-942.

Bajari, Patrick, Chenghuan Sean Chu and Minjung Park. "An Empirical Model of Subprime Mortgage Default from 2000-2007", NBER Working Paper no. 14625, December 2008.

Bayer, Patrick, Christopher Geissler and James Roberts. "Speculators and Middlemen: The Role of Flippers in the Housing Market", NBER Working Paper No. 16784, February 2011.

Bhutta, Neil. "Regression Discontinuity Estimates of the Effects of the GSE Act of 1992", *Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C.*, March 2009.

Brueckner, Jan, Paul Calem and Leonard Nakamura. "Subprime Mortgages and the Housing Bubble", working paper, 2011 (SSRN).

Bubb, Ryan and Alex Kaufman. "Securitization and Moral Hazard: Evidence from a Lender Cut-Off Rule", working paper, 2009 (SSRN).

Burnside, Craig, Martin Eichenbaum and Sergio Rebelo. "Understanding Booms and Busts in Housing Markets", NBER Working Paper No. 16734, January 2011.

Campbell, John, Stefano Giglio and Parag Pathak. "Forced Sales and House Prices", *American Economic Review*, forthcoming.

Card, David, Alexandre Mas and Jesse Rothstein. "Tipping and the Dynamics of Segregation", *Quarterly Journal of Economics*, Vol. 123, no. 1 (2008): 177-218.

Case, Karl and Robert Shiller. "Prices of Single Family Homes Since 1970: New Indexes for Four Cities", *New England Economic Review* (1987): 46-56.

_____, "The Efficiency of the Market for Single Family Homes", *American Economic Review*, Vol 70, no. 1 (1989): 125-137.

Chomsisengphet, Souphala and Anthony Pennington-Cross. "Evolution of the Subprime Mortgage Market", *Federal Reserve Bank of St. Louis Review*, Vol. 88, no. 1 (2006): 31-56.

Cutler, David M. James M. Poterba and Lawrence H. Summers. "Speculative Dynamics and the Role of Feedback Traders," NBER Working Papers 3243, 1991.

Danis, Michelle A. and Pennington-Cross, Anthony. "The delinquency of subprime mortgages," *Journal of Economics and Business*, vol. 60 (2008): 67-90.

Demyanyk, Yuliya and Otto Van Hemert. "Understanding the Subprime Mortgage Crisis", *Review of Financial Studies*, forthcoming.

Favilukis, Jack, Sydney Ludvigson and Stijn Van Nieuwerburgh. "Macroeconomic Implications of Housing Wealth, Housing Finance and Limited Risk Sharing in Equilibrium", NYU Stern Working Paper, January 2011.

Foote, Chris, Kris Gerardi and Paul Willen. "Negative Equity and the Foreclosure: Theory and Evidence", *Journal of Urban Economics*, Vol. 14, no. 2 (2008): 234-245.

Gerardi, Kristopher, Andreas Lehnert, Shane Sherlund and Paul Willen. "Making Sense of the Subprime Crisis", *Brookings Papers on Economic Activity*, Fall 2008: 69-145.

Gerardi, Kristopher, Adam Shapiro and Paul Willen. "Subprime Outcomes, Risky Mortgages, Homeownership Experiences and Foreclosures", Federal Reserve Bank of Boston Working Paper, 2007.

Glaeser, Edward and Gyourko, Joseph. "Urban Decline and Durable Housing", *Journal of Political Economy*, Vol 113, no 2, 345-375, April 2005

Glaeser, Edward and Joseph Gyourko. "Arbitrage in Housing Markets", Ch. 5 (pp. 113-146) in *Housing and the Built Environment: Access, Finance, Policy* (E. Glaeser and J. Quigley, eds.), Lincoln Land Institute of Land Policy, Cambridge, MA: 2010.

Glaeser, Edward, Joseph Gyourko, Eduardo Morales and Charles Nathanson. "Housing Dynamics", March 2011 (update of NBER Working Paper 13704).

Glaeser, Edward, Joshua Gottlieb, and Joseph Gyourko. "Can Cheap Credit Explain the Housing Boom?", NBER Working Paper No. 16230, July 2010.

Glaeser, Edward, Joseph Gyourko, Eduardo Morales and Charles Nathanson. "Housing Dynamics", June 2011 (update of NBER Working Paper 13704).

Goetzmann, William, Liang Peng and Jacqueline Yen. "The Subprime Crisis and House Price Appreciation", NBER Working Paper no. 15334, September 2009.

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

Gyourko, Joseph and Albert Saiz. "Construction Costs and the Supply of Housing Structure", *Journal of Regional Science*, Vol 46, no. 4 (November 2006): 661-680.

Hansen, Bruce. "Sample Splitting and Threshold Estimation", *Econometrica*, Vol. 68, no. 3 (May 2000): 575-603.

Haughwout, Andrew, Ebiere Okah and Joseph Tracy. "Second Chance: Subprime Mortgage Modification and Re-Default", *Federal Reserve Bank of New York Staff Reports*, report no. 417, Revised August 2010.

Himmelberg, Charles, Chris Mayer and Todd Sinai. "Assessing High House Prices: Bubbles, Fundamentals, and Misperceptions", *Journal of Economic Perspectives*, Vol. 19, no. 4 (Fall 2005): 67-92.

Ho, Giang and Anthony Pennington-Cross. "Predatory Lending Laws and the Cost of Credit", *Real Estate Economics*, Vol 36, no. 2 (2008): 175-211.

Keys, Benjamin, Tanmoy Mukherjee, Amit Seru and Vikrant Vig. "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans", *Quarterly Journal of Economics*, Vol. 125, no. 1 (February 2010): 307-362.

Krainer, John. "A Theory of Liquidity in Residential Real Estate Markets", *Journal of Urban Economics*, Vol. 49 (2001): 32-53.

Lai, Rose and Robert Van Order. "Momentum and House Price Growth in the United States: Anatomy of a Bubble", *Real Estate Economics*, Vol. 38, no. 4 (2010): 753-773.

Leamer, Edward. "Let's Take the Con Out of Econometrics", *American Economic Review*, Vol. 73, no. 1 (1983): 31-43.

Mayer, Christopher and Karen Pence. "Subprime Mortgages: What, Where and to Whom?", Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C., No. 2008-29 (2008).

Mayer, Christopher, Karen Pence and Shane Sherlund. "The Rise of Mortgage Defaults", *Journal of Economic Perspectives*, Vol. 23, no. 1 (Winter 2009): 27-50.

Mian, Atif and Amir Sufi. "The Consequences of Mortgage Expansion: Evidence from the U.S. Mortgage Default Crisis", *Quarterly Journal of Economics*, Vol 124, no. 4 (November 2009): 1449-96.

Mian, Atif, Amir Sufi and Francesco Trebbi. "The Political Economy of the U.S. Mortgage Default Crisis", *American Economic Review*, Vol.100, no. 6 (December 2010): 1967-1998.

Piazzesi, Monika and Martin Schneider. "Momentum Traders in the Housing Market: Survey Evidence and a Search Model", *American Economic Review*, Vol. 99, no. 2 (May 2009): 406-411.

Piskorski, Tomacz, Amit Seru and Vikrant Vig. "Securitization and Distressed Loan Renegotiation: Evidence from the Subprime Mortgage Crisis", Chicago Booth School of Business Research Paper No. 09-02, September 2009.

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

Poterba, James. "Tax Subsidies to Owner-Occupied Housing: An Asset Market Approach", Quarterly Journal of Economics, vol. 99 (1984), p. 729-745.

Quigley, John. "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: 1979.

Roback, Jennifer (1982). "Wages, Rents, and the Quality of Life", Journal of Political Economy, Vol. 90, no. 4 (December 1982): 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 Univerity Press, 1979.

Rosen, Kenneth and Lawrence Smith. "The Price-Adjustment Process for Rental Housing and the Natural Vacancy Rate", American Economic Review, Vol. 73, no. 4 (1983): 779-86.

Saiz, Albert. "The Geographic Determinants of Housing Supply", *Quarterly Journal of Economics*, Vol. 125, no. 3 (August 2010): 1253-1296.

Shiller, Robert. *Irrational Exuberance*. Princeton, NJ: Princeton University Press, 2005 (2nd edition).

Stein, Jeremy. "Prices and Trading Volume in the Housing Market: A Model with Down-Payment Constraints", *Quarterly Journal of Economics*, Vol. 110 (1995): 379-406.

Van Nieuwerburgh, Stijn and Weill, Pierre-Olivier. "Why Has House Price Dispersion Gone Up?", *Review of Economic Studies,*vol. 77 (October 2010): 1567-1606.

Wheaton, William. "Vacancy, Search, and Prices in a Housing Market Matching Model", *Journal of Political Economy*, Vol. 98, no. 6 (1990): 1270-1292.

Wheaton, William and Gleb Nechayev. "The 1998-2005 Housing 'Bubble' and the Current 'Correction': What's Different This Time?",*Journal of Real Estate Research*, Vol. 30, no. 1 (2008): 1-26.

Figure 1: National House Price Cycles—Case/Shiller and FHFA Real Price Indexes
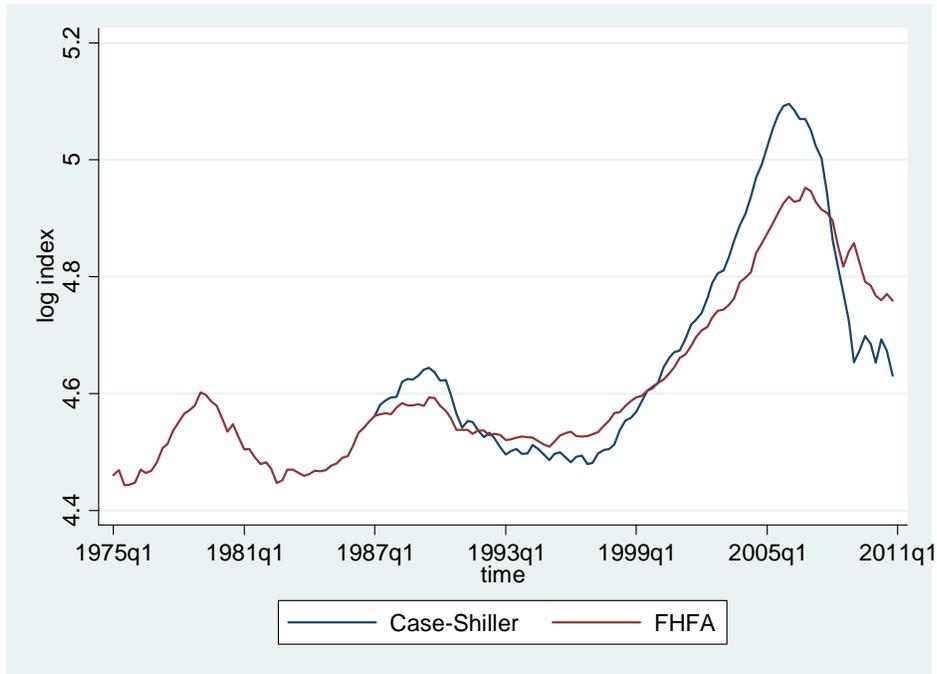

Figure 2: Hedonic and Repeat Sales Price Index Comparisons for Las Vegas
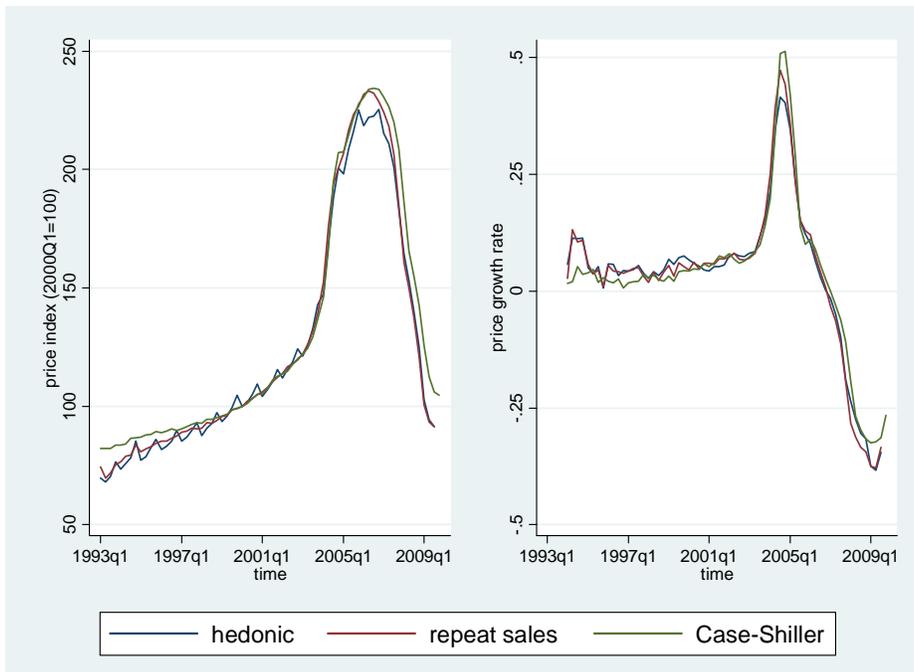
Notes: Hedonic price indexes were estimated as described in equation (3). The repeat sales index is based on a version of equation (3) that replaces house characteristics with a fixed effect for each house. That index uses a subsample of houses that had at least two transactions in DataQuick. The Case-Shiller index is publicly available.

Figure 3: Individual Metropolitan Area Hedonic House Price Indexes by Quarter
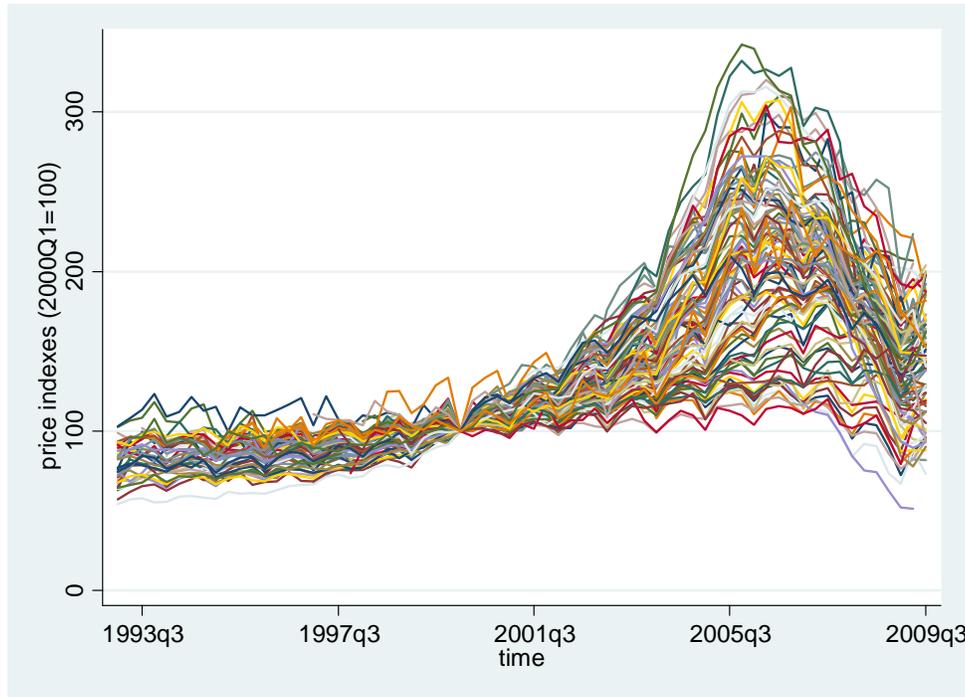
Notes: Each line represents a hedonic price index that was separately estimated for each MSA according to equation (3). The index for 2000Q1 is normalized to 100 for each MSA.

Figure 4: Neighborhood Hedonic Price Indexes, Selected MSAs
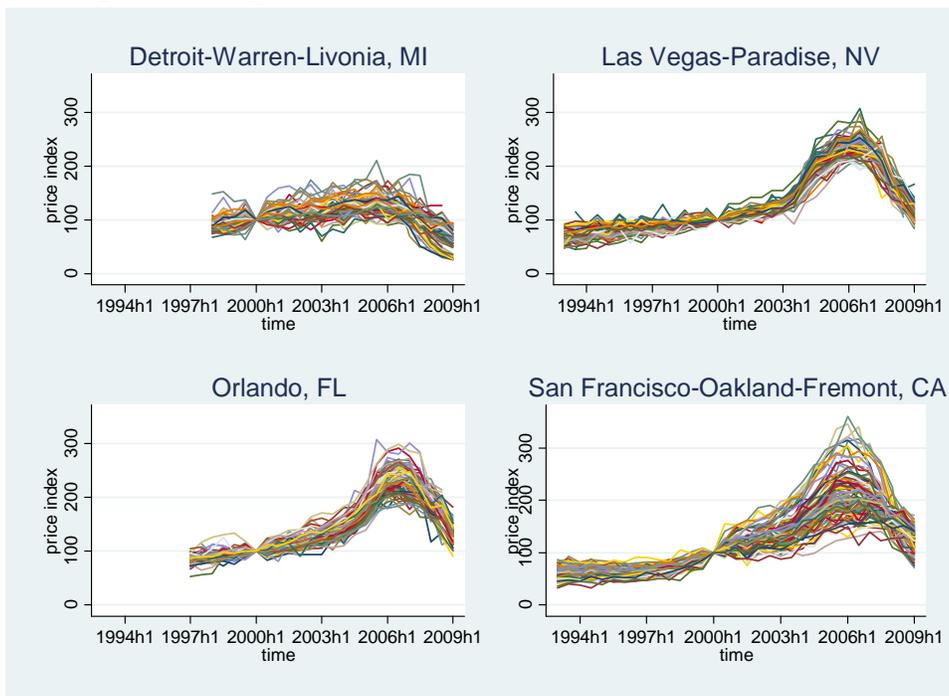
Notes: Each line represents a hedonic price index that was separately estimated according to equation (3) for each tract group within the relevant MSA. The index for 2000Q1 is normalized to 100 for each tract group.

Figure 5. Estimated Breakpoint Histogram, MSAs
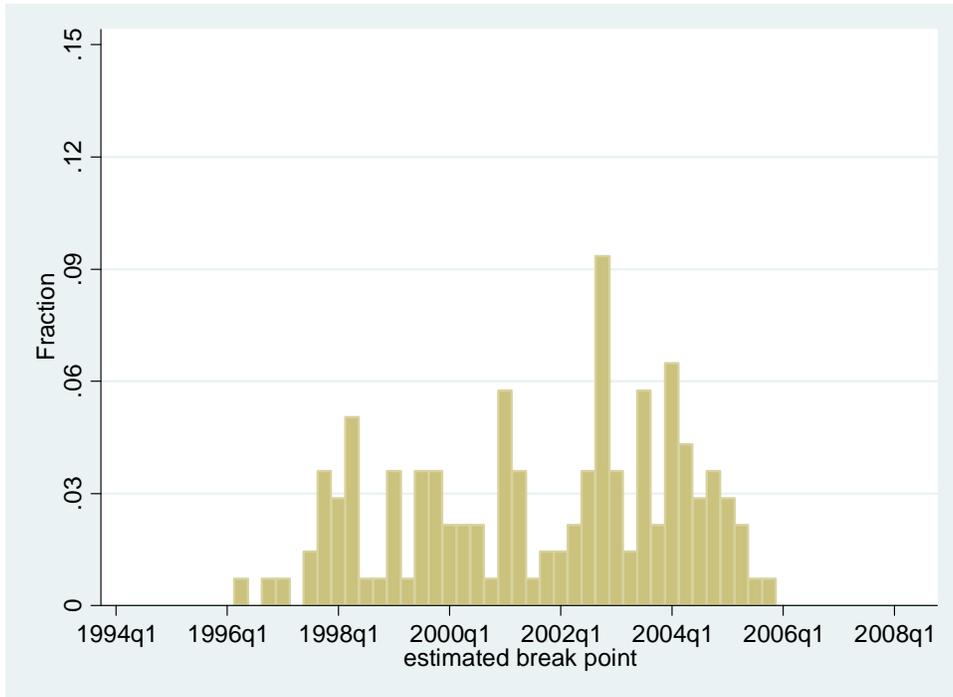
Notes: The histogram plots the fraction of estimated break points in each quarter using the sample of MSAs with statistically significant break points.

Figure 6. Estimated Breakpoint Histogram, Neighborhoods
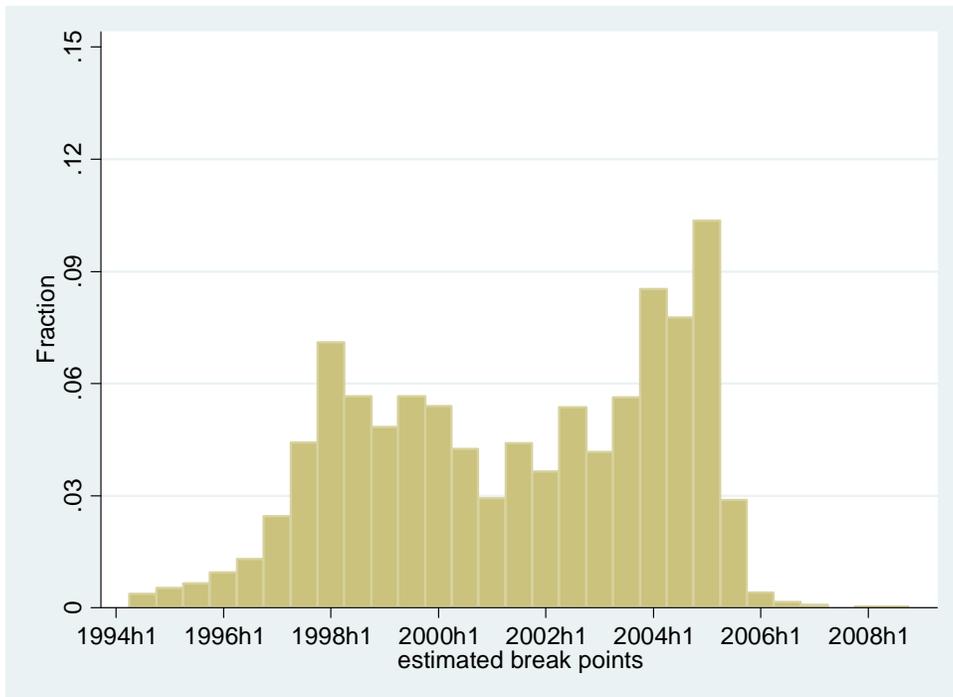
Notes: The histogram plots the fraction of estimated break points in each six month period using the sample of tract groups with statistically significant break points.

Figure 7. Estimated Breakpoints Histogram, Neighborhoods in Selected MSAs
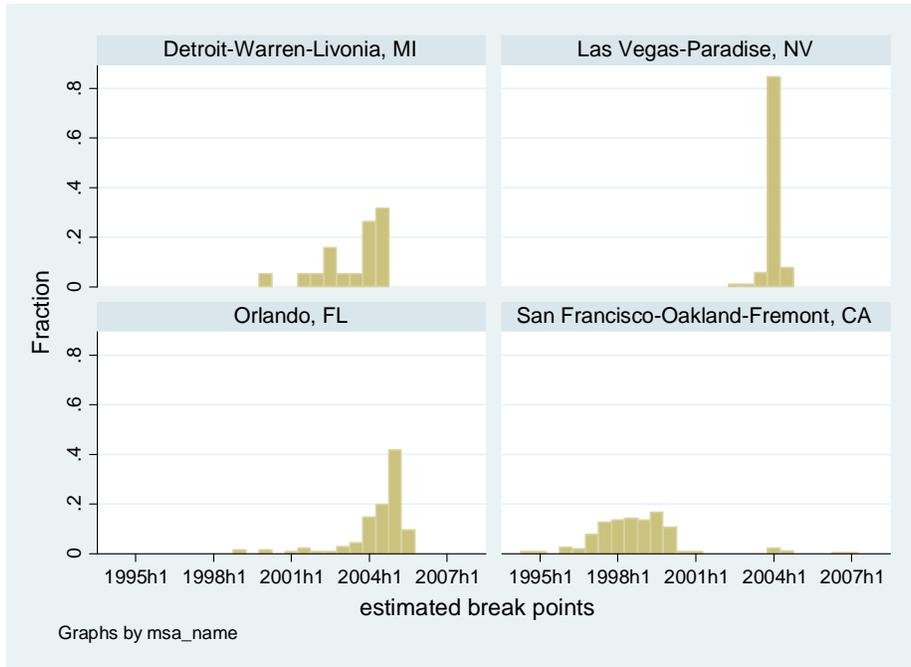
Notes: The histogram plots the fraction of estimated break points in each quarter using the sample of tract groups with statistically significant break points in each of the listed MSAs.

Figure 8: Estimated Price Appreciation Rates by Year Relative to Breakpoint, MSAs
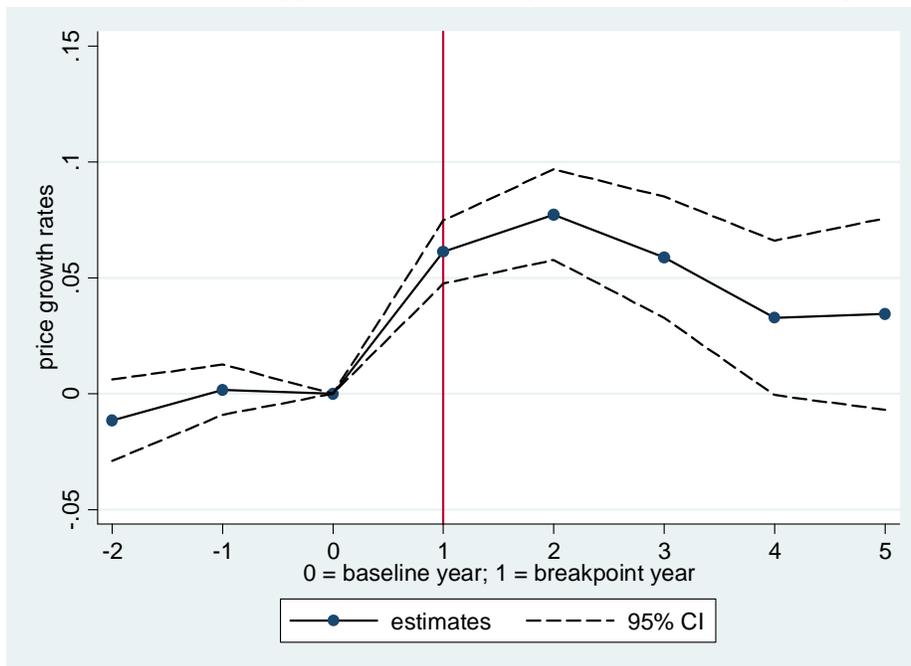
Notes: Each dot represents estimates of the magnitude of the price appreciation rates on an annual basis, relative to the estimated break point – see equation (5). Dashed lines are 95% confidence intervals. The sample includes the 79 MSAs that had statistically significant breakpoints. A random half-sample was used to estimate break points, with the other half used to estimate and plot the magnitudes.

Figure 9: Estimated Appreciation Rates Relative to Breakpoint, Selected MSAs
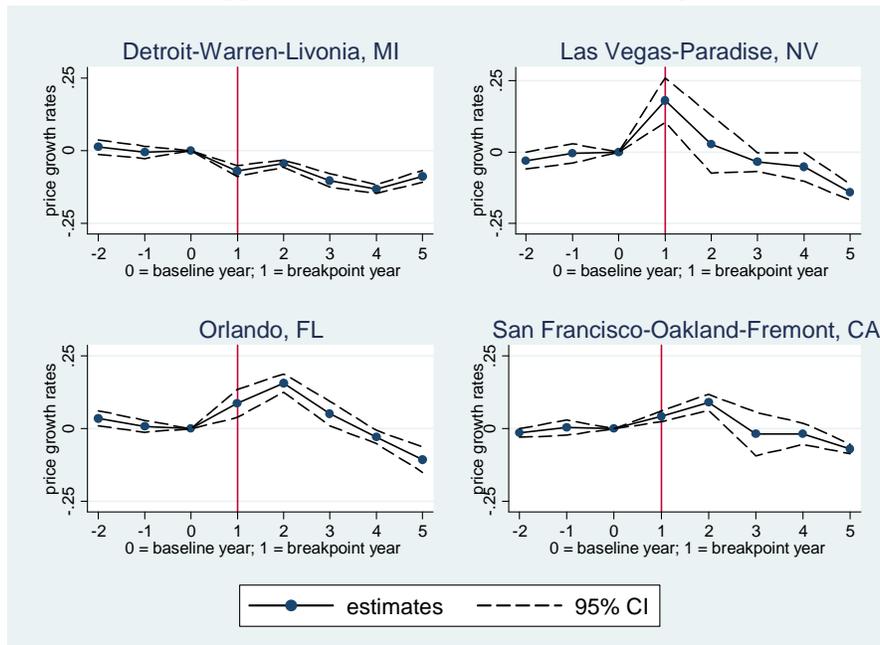
Notes: The point estimates in this figure come from separate regressions for each MSA. We first regress price growth data for all MSAs on quarter and MSA fixed effects, and then regress the residuals for each MSA on year dummies relative to the breakpoint year. Dashed lines are 95% confidence intervals. A random half-sample was used to estimate break points, with the other half used to estimate and plot the magnitudes.

Figure 10: Estimated Price Appreciation Rates by Year Relative to Breakpoint, Neighborhoods
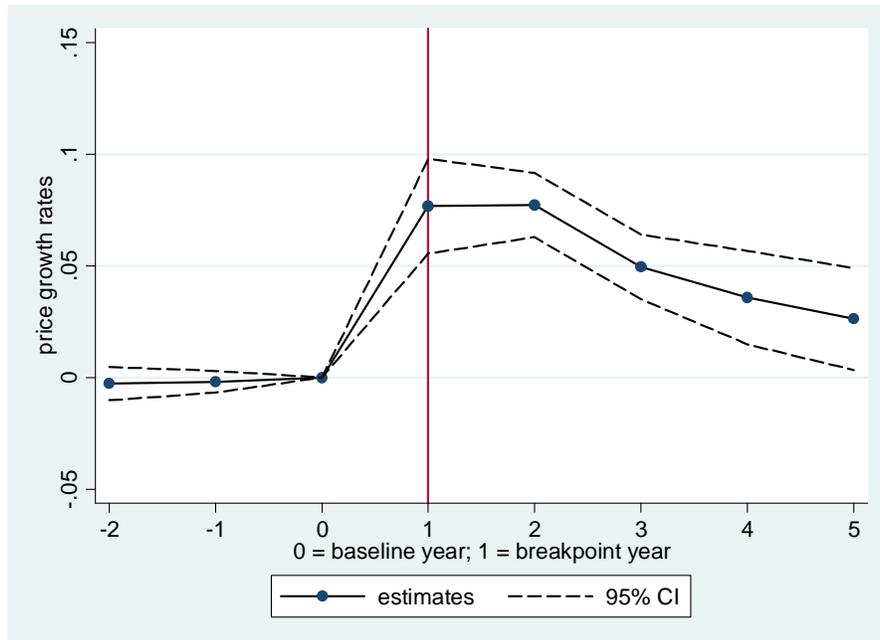
Notes: Each dot represents estimates of the magnitude of the price appreciation rates on an annual basis, relative to the estimated break point – see equation (5). Dashed lines are 95% confidence intervals. The sample includes 5,137 neighborhoods with statistically significant breakpoints. A random half-sample was used to estimate break points, with the other half used to estimate and plot the magnitudes.

Figure 11: Estimated Price Appreciation Rates Relative to Breakpoint, Neighborhood Data,
Selected MSAs
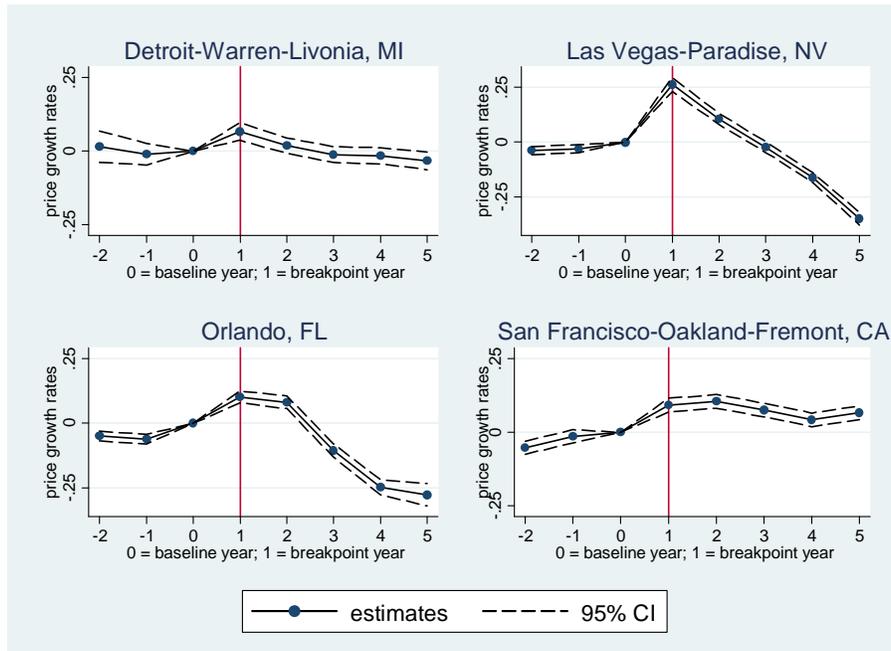

Notes: The point estimates in this figure come from separate regressions for each MSA. We first regress price growth data for all neighborhoods in the country on half-year and neighborhood fixed effects, and then regress the residuals for each neighborhood on year dummies relative to the breakpoint. Dashed lines are 95% confidence intervals. A random half-sample was used to estimate break points, with the other half used to estimate and plot the magnitudes.

Figure 12: Price and income comparisons relative to breakpoint, MSAs and Neighborhoods
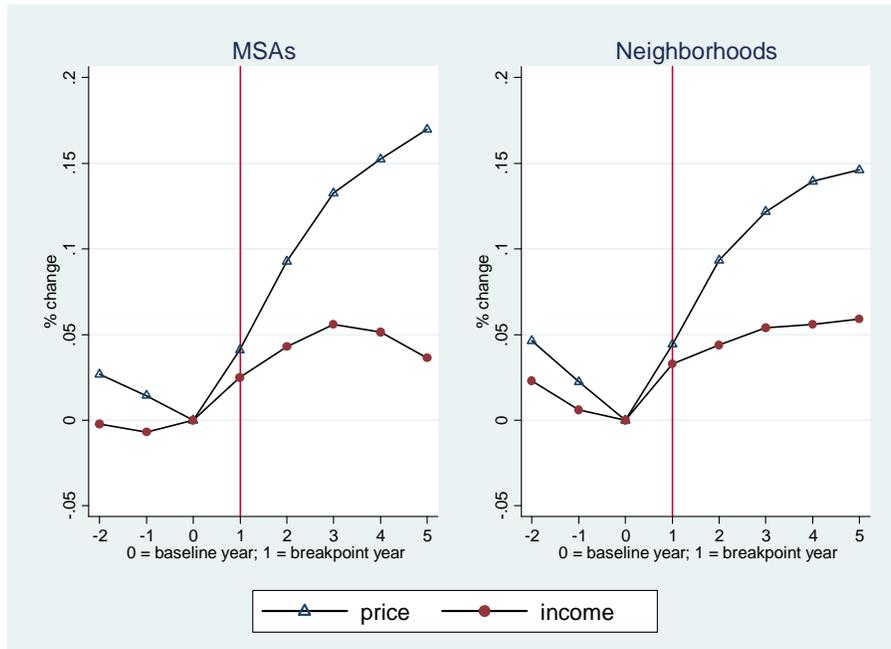
Notes: The point estimates for price and income correspond to the specification and results provided in Table 4. See the notes to that table for more detail.

.

Table 1: Summary Statistics on the Representativeness of our Final Sample

	All U.S.	DataQuick	Final
	(1)	(2)	(3)
Number of MSAs	362	269	94
Population of MSAs	642,486	809,386	1,322,485
	(1,485,668)	(1,691,640)	(2,520,843)
% East	0.21	0.22	0.24
	(.34)	(.35)	(.37)
% Midwest	0.22	0.20	0.17
	(.44)	(.41)	(.31)
% South	0.33	0.32	0.19
	(.49)	(.49)	(.44)
% West	0.24	0.26	0.40
	(.41)	(.44)	(.51)
% White	0.73	0.72	0.69
	(.14)	(.14)	(.15)
% Black	0.13	0.13	0.12
	(.11)	(.11)	(.09)
% College Degree	0.24	0.24	0.26
	(.07)	(.07)	(.07)
Median Family Income	$53,574	$54,017	$56,252
	(9,497)	(9,564)	(10,382)
Median House Value	$149,545	$153,381	$186,629
	(60,794)	(62,683)	(75,842)

Notes: All sociodemographic data based on Census 2000. First column presents averages and standard deviations (in parenthesis) for all MSAs in the country. Column 2 shows data only for MSAs covered by Dataquick. Column 3 presents descriptives for our final sample of 94 MSAs.

Table 2: Magnitude of the Breakpoint, MSAs and Neighborhoods

	dependent variable: price growth rate				
	(1)	(2)	(3)	(4)	(5)
MSA breakpoint	**0.074**	**0.054**	**0.056**	**0.065**	**0.061**
	(0.006)	**(0.005)**	**(0.006)**	**(0.007)**	**(0.007)**
Neighborhood breakpoint	**0.075**	**0.062**	**0.061**	**0.085**	**0.077**
	(0.010)	**(0.009)**	**(0.009)**	**(0.012)**	**(0.011)**
time effects	N	Y	Y	Y	Y
area effects	N	N	Y	Y	Y
dropping quarter of breakpoint	N	N	N	Y	N
areas with stat. sign. breaks	N	N	N	N	Y

Notes: This table presents the magnitude of the breakpoints for MSAs and Neighborhoods according to equation (5). Column 1 is not conditional on covariates, while column 2 adds time effects, and column 3 adds MSA fixed effects. Column 4 uses same specification as column 3, but drops the quarter of break point. Column 5 restricts the sample to MSAs with statistically significant breakpoints. Bold coefficients are significant at 5% level. MSA estimates use 4,956 observations, while neighborhood estimates use 94,833 observations.

Table 3: Prices and income around breakpoint

Dependent variable, estimates in logs	MSA averages (std dev) (1)	pre-trend		breakpoint jump		2 years after break	
		MSA (2)	neighborhood (3)	MSA (4)	neighborhood (5)	MSA (6)	neighborhood (7)
prices	118	-0.014	**-0.022**	**0.041**	**0.044**	**0.093**	**0.093**
	(24.63)	(0.009)	**(0.004)**	**(0.009)**	**(0.009)**	**(0.014)**	**(0.012)**
income	76881	0.007	-0.006	**0.025**	**0.033**	**0.043**	**0.044**
	(18423)	(0.008)	(0.005)	**(0.008)**	**(0.006)**	**(0.012)**	**(0.009)**

Notes: This table presents estimates of changes in the relevant variables around the breakpoint for both MSAs and neighborhoods using equation (5). Column 1 shows MSAs averages and standard deviations. All other estimates are in logs. Pre-trend columns show estimates for the 12-month period prior to the baseline, which is the 12-month period prior to the breakpoint. "breakpoint jump" columns correspond to the 12-month period starting with the quarter of the break, while the last two columns correspond to 2 years after the breakpoint. Bold coefficients are significant at 5% level.

Table 4: Heterogeneity in magnitude of price and income changes

	pre-trend		breakpoint jump		2 years after break	
	bottom 1/3 (1)	top 1/3 (2)	bottom 1/3 (3)	top 1/3 (4)	bottom 1/3 (5)	top 1/3 (6)
a) Timing						
MSA log prices	0.004	**-0.034**	**0.030**	**0.038**	**0.069**	**0.092**
	(0.013)	**(0.010)**	**(0.013)**	**(0.015)**	**(0.017)**	**(0.021)**
MSA log income	0.008	0.010	**0.027**	0.014	**0.033**	0.025
	(0.014)	(0.011)	**(0.011)**	(0.011)	**(0.014)**	(0.016)
neighborhood log prices	**-0.019**	**-0.029**	**0.021**	**0.066**	**0.050**	**0.143**
	(0.005)	**(0.011)**	**(0.006)**	**(0.017)**	**(0.011)**	**(0.023)**
neighborhood log income	-0.015	-0.003	**0.022**	**0.038**	**0.039**	**0.047**
	(0.008)	(0.008)	**(0.005)**	**(0.009)**	**(0.009)**	**(0.013)**
b) Supply elasticity						
MSA log prices	-0.015	-0.020	**0.038**	0.023	**0.082**	**0.053**
	(0.012)	(0.016)	**(0.014)**	(0.017)	**(0.015)**	**(0.016)**
MSA log income	-0.018	0.009	**0.025**	0.007	**0.038**	0.009
	(0.014)	(0.014)	**(0.010)**	(0.014)	**(0.011)**	(0.014)

Notes: This table presents heterogeneity in estimates in the relevant variables around the breakpoint for MSAs using a version of equation (5). It interacts all relative year dummies with dummies for bottom, middle, and top one-third parts of the distribution of breakpoint timing and supply elasticity (in separate models). All estimates are in logs. Pre-trend columns show estimates for the 12-month period prior to the baseline, which is the 12-month period prior to the breakpoint. "Breakpoint jump" columns correspond to the 12-month period starting with the quarter of the break, while the last two columns correspond to 2 years after the breakpoint. Bold coefficients are significant at 5% level.

Table 5: Other demand shifters and robustness tests around breakpoint

	MSA averages (std dev) (1)	pre-trend		breakpoint jump		2 years after break	
		MSA (2)	neighborhood (3)	MSA (4)	neighborhood (5)	MSA (6)	neighborhood (7)
Buyer Characteristics							
% minority	0.13	-0.003	**-0.003**	-0.003	**-0.008**	0.000	-0.006
	(0.10)	(0.002)	**(0.001)**	(0.002)	**(0.002)**	(0.004)	(0.003)
% white	0.72	0.003	**0.004**	0.001	**0.004**	-0.007	-0.003
	(0.13)	(0.003)	**(0.001)**	(0.003)	**(0.002)**	(0.005)	(0.003)
% speculation	0.12	0.002	**0.006**	0.003	**0.007**	0.006	**0.006**
	(0.03)	(0.002)	**(0.002)**	(0.002)	**(0.002)**	(0.003)	**(0.003)**
Credit Markets							
% subprime	0.15	-0.001	-0.003	0.002	0.000	0.009	**0.008**
	(0.06)	(0.004)	(0.002)	(0.004)	(0.001)	(0.006)	**(0.003)**
% FHA loans	0.12	-0.006	**-0.007**	**-0.010**	**-0.011**	**-0.018**	**-0.018**
	(0.10)	(0.005)	**(0.002)**	**(0.005)**	**(0.002)**	**(0.008)**	**(0.004)**
mean LTV	0.72	-0.005	**-0.006**	-0.001	**-0.005**	0.001	-0.005
	(0.12)	(0.004)	**(0.002)**	(0.004)	**(0.002)**	(0.006)	(0.004)
mortgage rate	6.58	**-0.338**	**-0.237**	**-0.141**	**-0.129**	**-0.109**	**-0.135**
	(0.82)	**(0.042)**	**(0.028)**	**(0.043)**	**(0.031)**	**(0.043)**	**(0.068)**
Prices and income							
rents	805	-0.003		0.014		0.025	
	(333)	(0.010)		(0.011)		(0.016)	
homebuyer income	78439	0.010	-0.003	**0.030**	**0.032**	**0.051**	**0.046**
	(18263)	(0.008)	(0.005)	**(0.008)**	**(0.006)**	**(0.012)**	**(0.008)**
per capita income	30028	**0.005**		**0.009**		**0.012**	
	(5673)	**(0.002)**		**(0.003)**		**(0.004)**	
Quantitites and supply							
transactions	2320	**0.039**	**0.064**	0.026	**0.051**	0.012	**0.048**
	(3044)	**(0.019)**	**(0.022)**	(0.020)	**(0.011)**	(0.030)	**(0.017)**
net migration	0.65	**0.233**		0.044		**-0.185**	
	(1.21)	**(0.057)**		(0.060)		**(0.088)**	
new supply	0.82	0.024	0.016	0.030	-0.008	0.036	0.072
	(0.62)	(0.022)	(0.020)	(0.025)	(0.026)	(0.039)	(0.039)
building permits	0.34	0.013		-0.001		-0.002	
	(0.25)	(0.011)		(0.014)		(0.021)	
construction costs	56608	0.001		0.002		-0.002	
	(6991)	(0.003)		(0.004)		(0.005)	
square footage	1770	6.160	3.097	-5.971	-3.682	-14.270	-9.397
	(172)	(5.460)	(2.430)	(5.088)	(3.800)	(7.842)	(5.515)
bathrooms	2.13	0.005	0.005	-0.005	-0.005	-0.011	-0.012
	(0.24)	(0.006)	(0.003)	(0.006)	(0.005)	(0.010)	(0.007)

Notes: This table presents estimates of changes in the relevant variables around the breakpoint for both MSAs and neighborhoods using equation (5). Column 1 shows MSAs averages and standard deviations. All other estimates are in logs. Pre-trend columns show estimates for the 12-month period prior to the baseline, which is the 12-month period prior to the breakpoint. "Breakpoint jump" columns correspond to the 12-month period starting with the quarter of the break, while the last two columns correspond to 2 years after the breakpoint. Bold coefficients are significant at 5% level.