Are there Bubbles in the Recent Housing Markets?

Bill Huang

December 2016

Abstract

Starting from 2012, the US house price has appreciated by 34.1% in nominal value, the largest appreciation since the housing boom from 2000 to 2006. Many literature has found evidence for the existence of bubble in the previous boom. On the other hand, given that 2012 is only two years after the Great Recession, the house price is likely undervalued. It is hard to determine whether the strong appreciation observed recently contains a bubble. In this study, I use a panel of 50 MSAs/MSADs with house price and economic fundamentals data from 1991 to 2015. I estimate the expected house price index by using the fundamentals and compare it to the actual real house price index for testing the presence of bubble. My results suggest that there is strong evidence of bubbles in 12 MSAs/MSADs in my sample, and all of them are in coastal regions except for Austin, TX.

1 Introduction

Since 2012, house price in United States has appreciated by 24.5% in total in real term, which is 5.63% per year. The annual growth rate is even slightly higher than the growth rate in the housing boom from 2000 to 2006 (5.31% per year in the 6-year-period). Figure 1 shows that the nominal house price index in 2016 has exceeded the historical high in 2006. Moreover, the real price index seem to reach the historical high level as well if the appreciation will persist for another two years or so. The collapse of the housing price from 2007 to 2011 clearly indicate the presence of "bubble" in the housing boom from 2000 to 2006. The burst of bubble has caused catastrophic impact to both the housing market, national economy, and even international economy. The real house price dropped 28.3% in 2011 since its historical high in 2006 nationally. Moreover, according to the "Cost of the Crisis" by Better Markets, a pro-financial reform Wall Street watchdog,

> The financial crash and its fallout will ultimately cost the hardworking American people more than $20 trillion in lost gross domestic product (GDP). Those losses include historically high unemployment, underemployment, long-term unemployment, foreclosure, underwater mortgages, bankrupt businesses large and small, lost savings, deffered or denied retirements, educations cut short, and so much more [1].

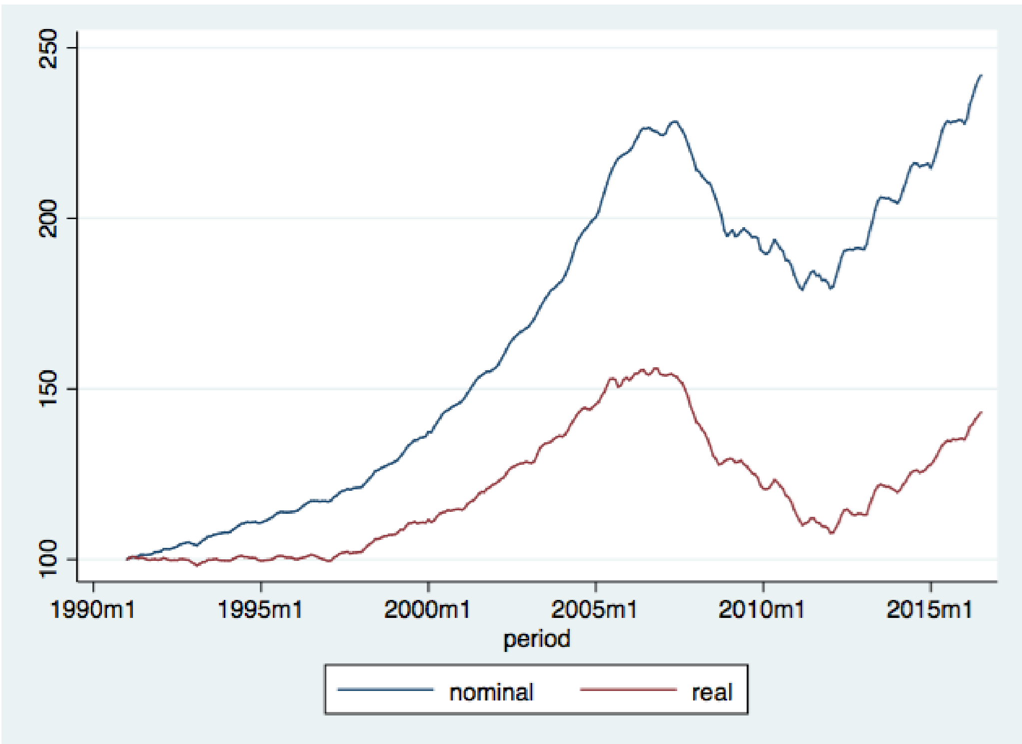


Figure 1: Nominal and Real House Price Index from 1991 January to 2016 July. Real term is calculated by Nominal value deflated by CPI less shelter. Source: FHFA, BLS.

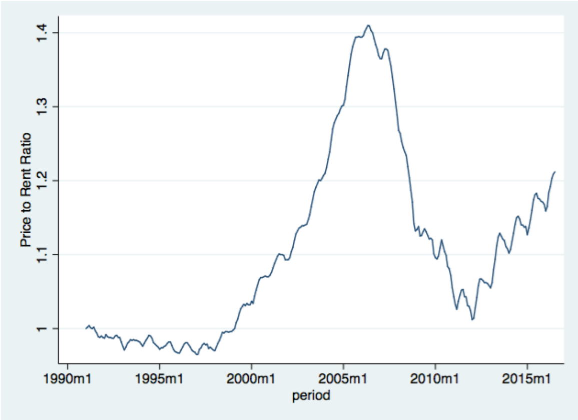


Figure 2: Ratio of Standard HPI to Owner's Equivalent Rent with base at 1991Q1. Source: FHFA, BLS.

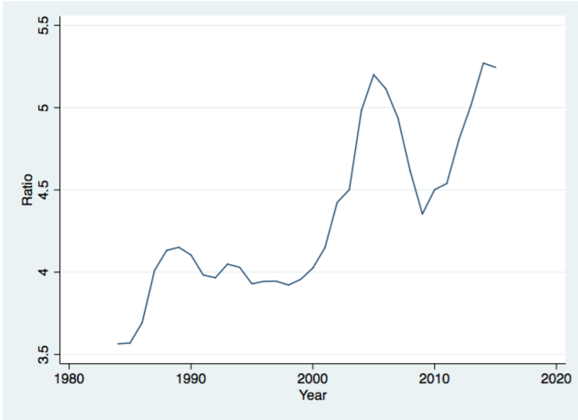


Figure 3: Ratio of Median Sales Prices of New Homes Sold to Median Household Income. Source: Census.

If we only focus on the cost of defaulting in mortgage itself, an analysis by the Economist suggests that since the America is nationalizing its mortgage market, the taxpayers are actually subsidising the housing borrowers. Another crisis will cost the taxpayer around 2-4% of GDP, which is not far off the cost of the 2008-09 bank bail out [2]. With such damaging losses the burst of housing bubble had done to the economy, one must be cautious about the potential of another bubble. However, we also need to consider the significant changes in the fundamental economic determinants of house prices to justify the existence of bubble. The appreciation can be resulted from the growth in population, increase in household income, and other related factors.

Figure 2 is the house price to rent ratio. The ratio rose dramatically since the beginning of the housing boom in the first decade, meaning that asset price appreciated relatively much faster than the rent price. Since 2012, the ratio rose again similar to the previous boom but with more fluctuation. This indicate the potential of bubble in the recent housing market. Figure 3 is the house price to income ratio, which is frequently used as a measure of home-ownership affordability. If the ratio is high, then the general household will find the house more expensive to afford, reducing the demand in the market and leading to decline in house price. Figure 3 shows that the ratio has exceeded its historical high recently, meaning that house is not as affordable as compared to the peak of previous bubble. This is a strong evidence for potential bubble in the recent housing market. However, this two ratio do not take into the change in mortgage rate and other fundamental factors, so it might lead to incorrect conclusion as McCarthy and Peach [10] suggested.

If the demand is high while the supply is low in the market, the appreciation in price is reasonable. If the demand is low while the supply is abundant, then we would expect the depreciation in housing price. Figures 4 and 5 shows the demand (proxy by home-

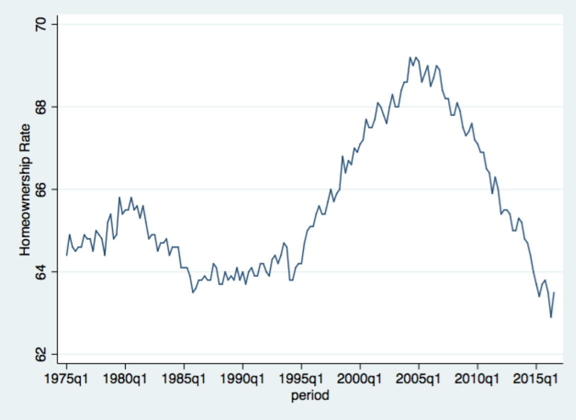


Figure 4: Home-ownership Rate. Source: Census.

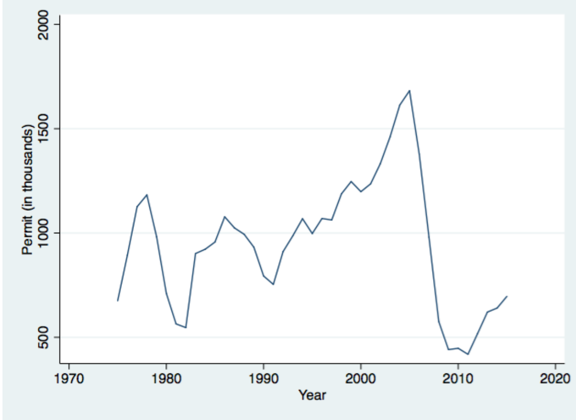


Figure 5: Single-Family Building Permits (in thousands). Source: Census.

ownership rate) and supply (proxy by single-family building permit [3]) from 1975 to 2016. In the housing boom from 2000 to 2006, home-ownership rate increase dramatically, indicating a strong demand in the housing market. However, over the same period, more building permits were issued than ever in the history, trying to satisfy the demand. When the supply finally caught up with the demand, the price become unsustainable, and the bubble burst. However, the demand for housing is decreasing in the recent market while the residential developments are also relatively inactive. Given that the market is somewhat stagnant recently, the appreciation in price is unexpected. Hence, more sophisticated analysis is needed to answer the question: are there housing bubbles?

Section 2 reviews the literature that is relevant to my study. Section 3 describes the theoretical model. Section 4 presents the data and the empirical model, and Section 5 the empirical results. Section 6 discusses the method of testing for bubbles. My conclusions are in Section 7.

2 Literature Review

After having risen significantly for around half of the decade, the housing price plummeted since the 2006 across the US. The dramatic up and down of the housing price in the first decade of the century has been the interest for many literature to investigate whether there were bubbles and when would the bubbles break down on national, regional, state, and metropolitan statistical areas (MSA) level housing markets. Different econometric techniques have been applied in testing the existence of the bubbles, but first, let me define what a bubble is in the real estate market. In the "Symposium on bubbles", Stiglitz (1990) provides a general definition for bubble:

> If the reason that the price is high today is only because investors believe that the selling price will be high tomorrow – when "fundamental" factors do not

> seem to justify such a price – then a bubble exists [13].

The above definition emphasized the importance of economic fundamental factors in the housing markets. On the demand side, demographic changes, income growth, the decline of mortgage rate, and the employment growth are the main factors in affecting the house prices. On the supply side, the attention has been paid to construction cost and regulation. If the actual house prices are systematically deviated from what the "fundamentals" predict, a "speculative bubble" might exist in the market. Moreover, Case and Shiller (2013) provides a intuition on why a bubble exists, which they state that a "bubble" refers to

> a situation in which excessive public expectations of future price increases cause prices to be temporarily elevated. During a housing price bubble, homebuyers think that a home that they would normally consider too expensive for them is now an acceptable purchase because they will be compensated by significant further value of their home to do the saving for them. First-time homebuyers may also worry during a housing bubble that if they do not buy now, they will not be able to afford a home later. Furthermore, the expectation of large price increases may have a strong impact on demand if people think that home prices are very unlikely to fall, and certainly not likely to fall for long, so that there is little perceived risk associated with an investment in a home [2].

However, when the price rises to a point that many people do not think it will continue to rise, the support for them to purchase a such expensive house will eventually break down. The diminishing demand in the market will cause the price to correct itself (depreciation) and return to the level justifiable by the economic "fundamentals" (the bubble bursts), and that is what we saw in housing market from the second half of 2006 to 2009 as indicated in many literature.

Many paper has discussed how economic fundamentals affect the local market in both long run and short run. Hwang and Quigley (2006) [9] investigate the effect of national and regional economic conditions on housing price, vacancies, and residential construction activity (building permits) of single-family housing market on a panel of 74 metropolitan statistical areas from 1987 to 1999. They use an simultaneous equation model consisted of three equations (Supply, Demand, and Vacancy). The empirical results confirm that house prices are strongly influenced by the the fundamentals such as income growth, employment, and regional economic conditions, and they also suggest that the local regulation influences how the market responds to the regional economic shocks. Furthermore, by looking into quarterly house price data from 1990 to 2009 (a period include the boom and the burst), Bhattacharya and Kim (2011) [1] still find compelling evidence that fundamental determinants (employment, construction cost, user cost) have a significant impact on housing price in the 20 metropolitan statistical areas in their study. Other literature build on the relationship between the house price and fundamental to determine whether there was a bubble in the beginning of the current century.

McCarthy and Peach (2004) [10] investigate the aggregate national house price from the mid-80s to 2003 by using a structural model that includes household income, user cost, and other fundamentals. Their results suggest that there was no bubble in the national housing market since mid-90s and the change in house price was attributable to the movement in personal income and nominal mortgage rate. Similar to the findings of McCarthy and Peach, Case and Shiller (2003) [2] conclude that there is no bubble on state-level home prices from 1985 to 2002 and the change in the fundamentals (income growth and falling interest rate) can explain much of the increase in the housing price. Moreover, Himmelberg, Mayer, and Sinai (2005) [7] use their own calculations to construct the local annual cost of owner-occupied housing for 46 metropolitan areas from 1980 to 2004, and they argue that the high price-to-income and price-to-rent ratios can be explained by the change in long-term interest rates and expected house price appreciation and therefore no evidence of bubble is found in the period. Though the above literature do not find evidences of bubble prior to 2004, mixed result is found in Goodman and Thibodeau (2008)'s study [5] that include data in 2005. They use a simultaneous equation model of demand and supply and estimate the long run effect of fundamental economic determinants on data in 1990 and 2000. With the estimated coefficients and supply elasticities, they calculate the expected growth rate for the 84 Metropolitan statistical areas (MSA) in the study and compare it with the actual growth rate: 39 MSAs experienced significantly more appreciation and they were concentrated in the northeast and west coast except for Las Vegas.

Different from the above literature, Mikhed and Zemčík [11], Clark and Coggin [3] test for a bubble from a time series approach. Other literature and the above two have tested that house price series and some economic fundamentals (e.g. income) have unit roots. If the house price and fundamentals (two unit root series) are cointegrated, then a linear combination of them is stationary, implying a equilibrium relationship (no bubble). However, if there is no cointegration, then there is no converge equilibrium relationship between the house price and fundamentals, implying the existence of bubble. Mikhed and Zemčík (2009) [11] cannot find any combination of non-stationary fundamentals that are cointegrated with the house price in a panel of 22 metropolitan statistical areas from 1978 and 2007 even after controlling for cross-sectional dependency in the cointegration test. Agreeing with Mikhed and Zemčík's study, Clark and Coggin (2011) [3] cannot discover the cointegration between the fundamentals and regional house price from 1975 to 2005 after allowing for structural break in the cointegration test. Both papers indicate the present of bubble in the housing market.

Lastly, Zhou and Sornette (2005) [14] test for a bubble from a more mathematical approach. They define a bubble as "faster-than-exponential" growth rate in price. By investigating national and state level data, they found that 22 states, mostly in Northeast and West, show a clear bubble from their definition. Moreover, they successfully predict that the bubble is unsustainable and the turning point of the bubble is likely to occur in mid-2006.

Many of the previous studies have looked into the period before 2006, but only a few cover the burst period (2006 - 2009) and there is still no literature concerned with the strong appreciation in the recent housing market. Having data from 1991 to 2015 provides me the chance to examine the potential of bubbles in the current housing markets.

3 The Theoretical Framework

This study considers the interaction of demand, supply, and vacancy effect in determining the house price level in equilibrium. From the demand perspective shown by equation (1), the demand for housing Q^D depends positively on population H, positively on household income Y, positively on local economic condition E, positively on rent price R, negatively on user cost UC, and the unobserved demand shifter ϵ^D.

$$Q^D = g_1(H, Y, E, R, \mathrm{UC}; \epsilon^D) \tag{1}$$

User cost represents the cost of owning a house. Following Goodman and Thibodeau's (2008) [5] and Bhattacharya and Kim's (2011) [1] studies, it depends positively on the price of the house P, positively on mortgage rate M, positively on income tax τ_y, positively on property tax τ_p, positively on depreciation rate δ, and negatively on expected capital gains g as shown in equation (2).

$$\mathrm{UC} = h(P, M, \tau_y, \tau_p, \delta, g) \tag{2}$$

From the supply perspective shown by equation (3), the supply for housing Q^S depends positively on house price P, negatively on structure cost C, negatively on land value L, negatively on regulation Reg, negatively on geographic constraint G, and the unobserved supply shifter ϵ^S

$$Q^S = g_2(P, C, L, \mathrm{Reg}, G; \epsilon^S) \tag{3}$$

The number of vacancy units V should depend on the quantity demand Q^D and quantity supply Q^S in the market. If there is more supply than demand, we expect the number of vacancy unit to be high. If the demand is strong relatively to the supply, then we will expect the number of vacancy unit to be low. Moreover, the number of vacancy unit depend positively on price P but negatively on economic condition E. If the price is high, then house become less affordable, which results in more vacancies. If economic condition is good, then the demand will be high so less vacancies. Lastly, when there is more uncertainty in price, VAR, home-owner might keep the home vacant until they are certain about the price trend before they make the decision of selling. However, it is also possible that the home-owner is risk averse so he sell the house in a short period of time. Hence, the unit of vacancy unit can be represented by equation (4).

$$V = g_3(P, Q^S, Q^D, E, \text{VAR}; \epsilon^V) \tag{4}$$

Different from markets for other goods, the equilibrium in the housing market is represented by equation (5).

$$Q^D = Q^S - V \tag{5}$$

From the above systems of equations, we can solve for the reduced form for P, which is

$$P = f(H, Y, E, R, M, \tau_y, \tau_p, \delta, g, C, L, \text{Reg}, G, \text{VAR}; \epsilon^D, \epsilon^S, \epsilon^V) \tag{6}$$

that include all the exogenous variables in the above equations.

4 Data and empirical specification

4.1 Data and descriptive statistics

The unit of observation in my study is the 50 largest Metropolitan Statistical Areas (MSA) and Division (MSAD) that have quarterly expanded-data Housing Price Indexes (HPI) [4] available on the FHFA (Federal Housing Finance Agency). The expended-data HPI is estimated using the same repeat transaction indexing methodology as is used in the construction of the standard HPI. Previous literature have raised concern about the accuracy of the standard HPIs since they are estimated only based on home value data from the Fannie Mae and Freddie Mac. It is limited to homes purchased with conventional mortgage loans or below the conforming loan limit [10]. The "expended-data" HPIs include some externally-sourced sales price data from county recorder offices and Federal Housing Administration (FHA)-endorsed loans, which are more reliable indicators for the movement of single-family house price in the MSAs/MSADs. The "expended-data' HPIs are available starting at 1991, so my study cover the period from 1991 to 2015.

The annual population and median household income on county level are available from the Census Bureau [5]. Since the 50 MSAs/MSADs available for the HPI are based on the revised Metropolitan Stataistcal Areas (MSAs) and Divisions as defined by the Office of Management (OMB) in February 2013 [6], the annual population estimates on MSA level constructed by researchers in Texas A&M University is not suitable due to the difference in delineations. Moreover, median household income on MSA level is not public available online. Hence, I decided to construct the annual population estimate and median household income from the county level data based on the 2013 statistical area delineation file. To get the estimated population for a MSA/MSAD, I take the aggregate sum for the estimated population of all counties in that MSA/MSAD; to get the median household income, I take the arithmetic mean of the median household income for all counties in that MSA/MSAD.

The monthly employment and unemployment data are public available from the Bureau of Labor Statistics (BLS) for the 50 MSAs/MSADs in my sample. Annual CPI less shelter and owners' equivalent rent of primary residences on the four regions (Midwest, Northeast, South, West) and selected cored based statistical areas are both available from the BLS. I match the MSA/MSAD in my sample to its core based statistical areas if available. If not, I match it to its regional data. CPI less shelter is used to deflate the nominal data to real term, and the owners' equivalent rent of primary residences is a proxy for the movement in the rent market.

Annual conventional single-family mortgage rate on selected large metropolitan areas (27 areas) from 1978 to 2015 is available on FHFA. I match the MSA/MSAD in my sample to the metropolitan areas in the mortgage rate data if available. If not, I assigned it a value based on a weighted average of mortgage rates from the surrounding metropolitan areas within 300 miles that have available mortgage rate data since the mortgage is mostly a local market in the 1990s [7].

Since there is no public available information on value of lands that are primarily for residential development (some private companies might have, but it will be expensive), I use the annual average farm real estate value from the department of agriculture as a proxy for movement in land value. The farm real estate value is a measurement of the value of all land and building on farms [8]. Since the building on farms does not change much from year to year, the change in farm real estate value is good approximation for the change in agricultural land value. The farm real estate value data is on state level, so to get the land value in a MSA/MSAD, I take the arithmetic mean of the farm real estate values of the states that the MSA is part of. The data on construction cost is also not public available (RS Means has construction cost data, but it is expensive), but data on structure cost is available on state level and for 46 selected MSAs from the Lincoln Institute of Land Policy [9]. The structure costs they provided is the average replacement cost of the housing structure calculated based on the RS Means data by Davis and Palumbo (2007) [4]. The change in estimated structure cost in a MSA at a given year can reflect the change in construction labor fee and material cost in that MSA at that year, so I use it as a proxy for the movement in the construction cost.

The data on geographic land scarcity is available from Saiz (2010) [12] who estimated the proportion of land unsuitable for housing development based on geographic characteristics, e.g. water bodies, for 95 MSAs. The index for regulatory land development restriction is the Wharton Residential Land Use Regulatory Index (WRLURI) from Gyourko, Saiz and Summers (2008) [6]. The index is developed based on a 2005 survey that provides information on local land use control regulation. The index is designed to have a mean 0 and standard deviation 1 across the nation. A higher value indicates a more restrictive environment for residential development. Since both data are available in Table I of Saiz's paper [10], I assign the corresponding MSA indexes to the MSA/MSAD in my sample [11].

I have data on annual, quarterly, and monthly level. For consistency, I convert all my data into annual level. For quarterly data, I take the arithmetic mean of the four quarters' data in a given year to get the annual estimation; for monthly data, I also calculate the arithmetic mean of the twelve months' data to get the estimation [12]. For data that are in nominal term (HPI, household income, rent index, structure cost, land value), I deflate them by using CPI less shelter to get the real term. Then I have a panel of 50 MSAs with 25 years (1991–2015) of data for the study.

Table 1 provides the summary statistics on real housing price index deflated by CPI less shelter for the 50 MSAs/MSADs in my sample. Over the period from 1991 to 2015, the median mean annual growth for the 50 MSAs/MSADs in my sample is 1.43%. Coastal MSAs/MSADs experience the larger growth than other MSAs/MSADs throughout the sample period. In the housing boom from 2000 to 2006, MSAs/MSADs in the Midwest had the weakest growth while MSAs/MSADs in the West experienced substantiated appreciation. Several MSAs/MSADs even doubled their real house price in only 6 years. For the MSAs/MSADs that experienced strongest growth in the housing boom, they also suffered the largest loss when the bubble bursted at mid-2006. For the recent appreciation since 2012, the median growth rate is 25.71% for the 50 MSAs/MSADs in only three-year-period. If the house price can appreciate for another 17.3% in the next three years, then the median growth rate since 2012 will outperform the median growth rate 47.76% from 2000 to 2006 [13]. Given that there was a bubble in the housing boom from 2000 to 2006 and it did burst and cause a disaster in the housing market and economy, we need to concern about the existence of a bubble if the house prices in the recent market continue to appreciate strongly. However, the strongest appreciations in the recent markets are happening in MSAs/MSADs that suffered the biggest depreciation during the burst period. It is possible that the house prices are undervalued for those markets and the strong appreciation we observe currently is the readjustment for the house prices in the undervalued market to return to its expected values. If it is the later case, we might

Table 1: Real Housing Price Index Summary Statistics

MSA/MSAD	Mean Annual Change(%)	Standard Deviation	%Δ 00 − 06	%Δ 06 − 09	%Δ 12 − 15
Midwest					
Chicago-Naperville-Arlington Heights, IL	0.32%	7.45	43.30%	−35.22%	19.69%
Cincinnati, OH-KY-IN	−0.30%	4.31	2.68%	−21.55%	13.45%
Cleveland-Elyria, OH	−1.07%	6.09	−4.84%	−33.42%	14.06%
Columbus, OH	0.04%	4.77	1.01%	−21.31%	20.59%
Detroit-Dearborn-Livonia, MI	−0.98%	11.60	−18.77%	−57.85%	48.70%
Indianapolis-Carmel-Anderson, IN	−0.08%	3.41	−5.09%	−14.54%	15.70%
Kansas City, MO-KS	0.84%	4.42	12.71%	−13.97%	16.42%
Milwaukee-Waukesha-West Allis, WI	0.90%	5.16	32.60%	−21.66%	13.18%
Minneapolis-St. Paul-Bloomington, MN-WI	1.83%	7.47	39.04%	−33.43%	25.09%
St. Louis, MO-IL	0.16%	4.97	22.83%	−20.50%	13.90%
Warren-Troy-Farmington Hills, MI	0.25%	9.96	−4.98%	−50.45%	49.76%
Northeast					
Boston, MA	2.30%	7.04	46.30%	−21.84%	21.66%
Cambridge-Newton-Framingham, MA	2.18%	6.24	35.62%	−19.85%	20.52%
Montgomery County-Bucks County-Chester County, PA	0.63%	5.03	50.66%	−12.10%	7.52%
Nassau County-Suffolk County, NY	1.95%	7.83	83.90%	−27.13%	11.79%
New York-Jersey City-White Plains, NY-NJ	1.47%	7.32	80.23%	−24.2%	12.07%
Newark, NJ-PA	1.04%	7.17	70.40%	−27.41%	8.21%
Philadelphia, PA	1.01%	5.79	70.74%	−7.09%	7.06%
Pittsburgh, PA	0.55%	2.88	5.50%	−7.42%	12.41%
Providence-Warwick, RI-MA	1.26%	8.34	77.43%	−29.81%	17.55%
South					
Atlanta-Sandy Springs-Roswell, GA	0.19%	6.95	4.67%	−26.35%	37.55%
Austin-Round Rock, TX	3.06%	5.20	3.22%	0.19%	33.99%
Baltimore-Columbia-Towson, MD	0.72%	6.93	80.10%	−21.59%	0.93%
Charlotte-Concord-Gastonia, NC-SC	0.49%	4.52	4.89%	−13.39%	20.61%
Dallas-Plano-Irving, TX	1.23%	4.57	−0.08%	−7.86%	32.44%
Fort Lauderdale-Pompano Beach-Deerfield Beach, FL	2.08%	12.28	118.73%	−52.97%	39.41%
Fort Worth-Arlington, TX	0.93%	4.03	0.91%	−7.23%	26.32%
Houston-The Woodlands-Sugar Land, TX	1.54%	4.59	8.32%	−5.31%	29.78%
Jacksonville, FL	0.68%	8.15	58.17%	−35.92%	21.89%
Miami-Miami Beach-Kendall, FL	2.70%	12.77	124.21%	−52.41%	42.64%
Nashville-Davidson–Murfreesboro–Franklin, TN	1.52%	4.72	15.28%	−11.77%	24.54%
Orlando-Kissimmee-Sanford, FL	0.99%	11.43	84.94%	−49.54%	35.36%
San Antonio-New Braunfels, TX	1.49%	3.62	15.34%	−5.11%	20.21%
Tampa-St. Petersburg-Clearwater, FL	1.05%	10.07	77.63%	−45.08%	29.93%
Virginia Beach-Norfolk-Newport News, VA-NC	1.59%	7.23	84.18%	−14.75%	8.28%
Washington-Arlington-Alexandria, DC-VA-MD-WV	1.73%	9.70	106.94%	−40.26%	16.01%
West Palm Beach-Boca Raton-Delray Beach, FL	1.69%	11.82	104.98%	−48.44%	42.25%
West					
Anaheim-Santa Ana-Irvine, CA	2.77%	11.21	114.10%	−38.12%	32.85%
Denver-Aurora-Lakewood, CO	3.36%	6.40	5.99%	−19.49%	36.22%
Las Vegas-Henderson-Paradise, NV	1.00%	13.51	82.61%	−56.66%	49.25%
Los Angeles-Long Beach-Glendale, CA	2.34%	12.59	133.92%	−44.86%	43.27%
Oakland-Hayward-Berkeley, CA	2.55%	12.19	74.05%	−49.86%	48.25%
Phoenix-Mesa-Scottsdale, AZ	2.19%	12.46	76.94%	−53.09%	36.83%
Portland-Vancouver-Hillsboro, OR-WA	3.40%	7.09	49.21%	−19.44%	34.07%
Riverside-San Bernardino-Ontario, CA	1.66%	14.23	134.34%	−55.82%	47.96%
Sacramento–Roseville–Arden-Arcade, CA	1.39%	12.71	101.63%	−49.43%	45.89%
San Diego-Carlsbad, CA	2.42%	11.32	94.15%	−41.92%	34.18%
San Francisco-Redwood City-South San Francisco, CA	3.28%	9.16	45.17%	−30.75%	42.51%
San Jose-Sunnyvale-Santa Clara, CA	2.89%	10.05	42.68%	−38.87%	37.30%
Seattle-Bellevue-Everett, WA	2.55%	7.64	49.24%	−21.45%	37.14%
Median	1.43%	7.28	47.76%	−26.74%	25.71%

Table 2: Demand and Supply Fundamentals Summary Statistics

Variable	Median	Mean	Stdev.	Min	Max
Demand Fundamentals					
Change in Population (%)	30.58	37.89	32.77	−16.55	156.26
Change in Real Household Income (%)	16.72	15.74	10.94	−13.30	45.34
Average Unemployment Rate (%)	5.64	5.83	0.95	4.21	8.64
Average Effective Mortgage Rate (%)	6.34	6.29	0.13	5.98	6.45
Change in Real Rent (%)	18.50	16.03	9.69	−9.94	34.22
Supply Fundamentals					
Change in Real Structure Cost (%)	52.00	52.12	28.55	−20.73	116.93
Change in Real Land Value (%)	128.96	125.63	78.85	34.16	466.72
Undevelopable Area (%)	30.26	30.21	21.90	1.44	76.63
Wharton Regulation Index	0.31	0.31	0.65	−0.79	1.89

not need to worry about the housing bubble. Moreover, it also indicate that our housing market is becoming healthier. Detroit [14] is one of the example.

Table 2 provides the summary statistics for the demand and supply fundamentals I am using in the study. For the median of the 50 MSAs/MSADs in my sample, population, real household income, real rent, real structure cost, and real land value all show significant growth over time. This may explain some of the growth in house price over the 25-year-period. The average unemployment rate and average effective mortgage rate are similar across the MSAs/MSADs in my sample. However, the MSAs/MSADs are very different geographically. Indianapolis-Carmel-Anderson, IN has the smallest proportion of land that is unsuitable for housing development (1.44%), while Miami-Miami Beach-Kendall, FL MSAD has 76.63% of its land unsuitable for residential development. The average Wharton Regulation Index is 0.31 and the standard deviation is 0.65, meaning that the MSAs/MSADs in my sample are more restrictive and homogeneous in regulation control than the average US cities [15].

4.2 The empirical specification

Since many literature (Mikhed and Zemčík [11], Clark and Coggin [3]) have shown that the real house price series and some economic fundamental determinant (real income) series are unit roots, I will run into the issue of spurious regression if I regress house price index directly on the fundamentals that have unit roots. Moreover, the house price data is a index while the income, real structure cost, and some other variables are in real value. It also does not make much sense in interpretation if I regress house price index directly on the fundamentals. What I decide to do is to transform my variables into either annual percentage change or annual difference. Hence, equation (6) becomes

$$\begin{aligned}\%\Delta P_{i,t} = \beta_1 &+ \beta_2\%\Delta H_{i,t} + \beta_3\%\Delta Y_{i,t} + \beta_4\%\Delta R_{i,t-1} \\ &+ \beta_5\Delta M_{i,t} + \beta_6\Delta \mathrm{UR}_{i,t} + \beta_7\%\Delta C_{i,t} \\ &+ \beta_8\%\Delta L_{i,t} + \beta_9\mathrm{VAR}_{i,t-1} + u_{i,t}\end{aligned} \tag{7}$$

where i stands for MSA/MSAD i and t for year. For variable $X_{i,t}$, $\%\Delta X_{i,t} = (X_{i,t} - X_{i,t-1})/X_{i,t-1} \times 100$ and $\Delta X_{i,t} = X_{i,t} - X_{i,t-1}$, so t goes from 1992 to 2015 in equation (7). $\%\Delta P_{i,t}$ is the annual percentage change in real price; $\%\Delta H_{i,t}$ is the annual percentage change in population; $\%\Delta Y_{i,t}$ is the annual percentage change in real household income; $\%\Delta R_{i,t-1}$ is lag annual percentage change in real rent, the reason I use one year lag here is that people need to wait toward the end of their rent term (generally one or two years, but some might be longer) to make their decision of whether purchasing a house or starting a new rent; $\Delta M_{i,t}$ is the annual difference in mortgage rate; $\Delta\mathrm{UR}_{i,t}$ is the annual difference in unemployment rate, which I use as a indicator for local economic condition E in equation (6); $\%\Delta C_{i,t}$ is the annual percentage change in real construction cost; $\%\Delta L_{i,t}$ is the annual percentage change in land value; $\mathrm{VAR}_{i,t-1}$ is the one lag uncertainty in price, the reason I use one year lag is that people do not observe the uncertainty at the market now so they will use the uncertainty in previous year as a indicator for the uncertainty in the market now [16]. Moreover, the MSA/MSAD fixed effect is taking into account once I take the percentage change as my dependent variable, so geographic constraint G that is fixed over time get dropped out from the model. There is generally very small change in a MSA/MSAD residential regulation over years. MSA/MSAD with stricter regulation rule will tend to be more restrictive in residential development in the future as compared to other MSAs/MSADs. We will make a assumption that the restriction in residential development do not change over time for the MSAs/MSADs in my sample period, then we can drop it from the model after we take the percentage change. I cannot find data on income tax rate τ_y, property tax rate τ_p, and depreciation rate δ. I assume that their changes are small over year, so by excluding them from the model it should not make dramatically difference on my prediction for expected growth later when testing for bubble. Last, I do not include the expected capital gains g in equation (7), either [17].

To construct the estimation for uncertainty in price for year t, I take the quarterly nominal (people only observe nominal value) house price data in year $t-1$ and fit a straight line for the four points. One fourth of the sum of square for the four residuals is the estimation for the uncertainty at year t. People tend to picture the housing price as a linear trend in one year period since it is the simplest model, then the noise around that trend represents the people's perception of fluctuation in the market. The bigger the fluctuation is, the larger the uncertainty is.

Because my sample contains a bubble period (2000 to 2006) and a burst period (2007-2009) which are captured by the error term $u_{i,t}$ in equation (7), we will get systematically positive residuals for the bubble period and systematically negative residuals for the burst period by estimating equation (7) in OLS. Moreover, the bubble is uncorrelated with other fundamentals I include in the model, so we can think of it as a random shock in time that is common to all MSAs/MSADs. The bubbles might have stronger impact in some cities (e.g. MSAs in the West), but the effect of bubbles might be alleviated by other factors in cities with weaker appreciation (e.g. some MSAs in the Midwest). Hence, my error term in equation (7) has two components: a time-specific random effect a_t ($a_t \sim \mathcal{N}(0, \sigma_t^2)$) and the white-noise-error $\epsilon_{i,t}$ ($\epsilon_{i,t} \sim \mathcal{N}(0, \sigma_\epsilon^2)$) that captures the normal fluctuation in the housing market in MSA i at time t. Equivalently, $u_{i,t} = a_t + \epsilon_{i,t}$ in equation (7). By doing a random effect model in the time dimension, we can get efficient estimate for the coefficients.

5 Empirical Results

5.1 Impact of Economic Fundamentals

Table 3 presents both the pooled OLS and year random effect regression results. From the discussion in the last section, the OLS estimator is inefficient. Moreover, the Breusch-Pagan Lagrange Multiplier test rejected the null hypothesis that there is no significant difference across year (no panel effect) at 1% significance level [18], suggesting that a random effect or fixed effect should be included in the model rather than using a pooled OLS. To include OLS estimators here is to provide a comparison, and the signs for all independent variables are the same across the two models. From the year random effect estimators, the impact of population change is positive and significant, which is in consensus with the hypothesis that more population cause higher demand for housing which leads to appreciation in price. Real household income has a positive and significant impact on the housing price as well, agreeing with the assumption that when people have higher income, they will buy better houses. The lag change in real rent also shows positive and significant impact, confirming the hypothesis that people have a stronger incentive to become homeowner when the rent is rising. The change in unemployment rate has a negative and significant impact on the housing price, which is in consensus with the idea that the depression in local economic leads to the inactive activity in housing market so price would drop. The change in effective mortgage interest rate also has a negative yet not significant (but significant in OLS) impact on the housing price, but the negative sign supports the idea that when people's burden of owning a house becomes low (mortgage is low), more people will want to purchase a house, driving the price in the market. Both change in real structure cost and change in real land value have positive impact on the

Table 3: Dependent Variable is Percentage Change in Real House Price Index

Variables	Pooled OLS	Year Random Effect
%Δ Population	0.416***	0.447**
	(0.152)	(0.182)
%Δ Real Median Household Income	0.485***	0.749***
	(0.077)	(0.141)
lag %Δ Real Rent	0.766***	0.536**
	(0.113)	(0.246)
Δ Unemployment Rate	−2.456***	−2.917***
	(0.261)	(0.585)
Δ Effective Mortgage Interest Rate	−1.217***	−1.646
	(0.308)	(1.343)
%Δ in Real Structure Cost	0.668***	0.231
	(0.105)	(0.162)
%Δ in Real Land Value	0.289***	0.211***
	(0.031)	(0.062)
lag Price Uncertainty	−0.444***	−0.198***
	(0.073)	(0.095)
Intercept	−1.347***	−1.084
	(0.379)	(0.999)
R^2	0.4436	0.4002
F	78.49***	
Wald χ^2_8		208.90***
σ_u		4.94
σ_e		4.30
N	1200	1200

Random Effect coefficients are estimated by GLS.
Robust standard errors in parenthesis. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Two-tailed test.

house price, agreeing with the supplier's psychology that a good should be sold in higher price when the raw materials become more expensive. Lastly, the lag price uncertainty has a negative but significant impact in the model, suggesting that home owners are willing to sell at a lower price when there was more fluctuation in the market. This behavior reflects that home owners are risk-averse. All the fundamentals in my model impact the housing price in the ways that I am expecting.

The above results indicate that the local economic fundamentals do play important roles in determining the house price in local market, confirming the findings of many literature listed in the literature review section. However, it is quite surprising that the change in effective mortgage interest rate is not significant in the year random effect model while it is significant in the pooled OLS. McCarthy and Peach (2004) [10], Case and Shiller (2003) [2], Battacharya and Kim (2011) [1] all claim that the change in nominal mortgage rate is strongly attributable to change in house price. The reason that the mortgage rate is not significant here is possibly that the time random effect might take up some of the effect caused by the change in mortgage rate.

Table 4: The impact of regulatory and geographic constraints.

Period	Undevelopable Area (%)	Wharton Regulation Index
Boom (00 - 06)	0.099^{***}	2.548^{***}
	(0.0211)	(0.369)
Bust (07 - 09)	-0.124^{***}	-0.578^{*}
	(0.036)	(0.304)
Other Years	0.013	0.277
	(0.009)	(0.203)

Dependent Variable is Percentage Change in Pirce
Robust standard errors in parenthesis. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Two-tailed test.

5.2 Regulatory and Geographic Constraints in the Boom and Burst

Though in the previous section I mentioned that the regulatory and geographic constraints that are presumably fixed over my sample period get dropped out from the model after I take the percentage change, Huang and Tang's study (2012) [8] concluded that both regulatory and geographic constraints contributed to the amplitude of price booms and bursts between 2000 and 2009. In another word, the regulation and geographic constraints additonally amplified the appreciation during the boom while they also make the depreciation more severed in the burst. I am going to test it by including the Wharton Regulation Index (WRI) and Proportion of Undevelopable areas (UA) with allowing different coefficients at different periods into equation (7). The periods are: the boom period from 2000 to 2006, the burst period from 2007 to 2009, and all other years. I use the same boom and burst period as in Huang and Tang's study. I run it with the year random effect model and the resulted coefficients on WRI and UA are presented in Table 4. The fundamentals show the similar impact on housing price as compared to the impact presented in Table 3, but the change in mortgage rate is significant at 10% level and the percentage change in real structure cost is significant at 1% level in this model. From Table 4, both UA and WRI have positive and significant impact on the housing price during the boom period, and they both show significantly negative impact on the price in the burst period. Moreover, in other years that are in my sample, both UA and WRI do not have significant impact on the housing price, indicating the additional impacts of UA and WRI are unique only in the boom and the burst periods. This strongly supports Huang and Tang's conclusion. Moreover, this suggests that areas with stricter regulation and more geographic constraints should be more concerned about the existence of a bubble since the appreciation will get amplified and the burst will be severer.

Table 5: Regional Structural Break?

Variable	West	Northeast	South	West
%Δ Population	$(0.00, 1.31)$	$(0.07, 1.39)$	$(0.06, 0.63)$	$(0.00, 0.73)$
%Δ Real Median Household Income	$(0.71, 1.18)$	$(0.24, 0.73)$	$(0.66, 1.03)$	$(0.19, 0.75)$
lag %Δ Real Rent	$(0.12, 0.81)$	$(0.48, 1.04)$	$(-0.10, 0.51)$	$(0.82, 1.49)$
Δ Unemployment Rate	$(-2.18, -1.07)$	$(-1.98, -0.46)$	$(-3.24, -1.71)$	$(-5.12, -3.54)$
Δ Effective Mortgage Rate	$(-3.46, 0.21)$	$(-4.00, -0.21)$	$(-3.58, 0.10)$	$(-4.12, -0.72)$
%Δ in Real Structure Cost	$(-0.072, 0.30)$	$(0.00, 0.66)$	$(0.05, 0.51)$	$(0.31, 1.04)$
%Δ in Real Land Value	$(-0.01, 0.22)$	$(0.19, 0.41)$	$(0.19, 0.32)$	$(-0.02, 0.28)$
lag Price Uncertainty	$(-0.17, 0.10)$	$(-0.17, 0.09)$	$(-0.37, 0.04)$	$(-0.43, -0.13)$

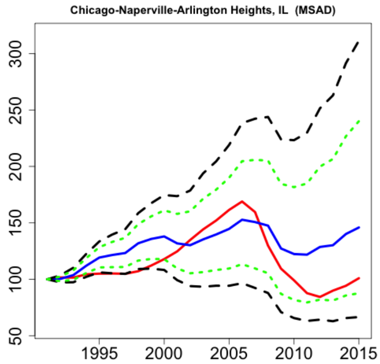
Dependent Variable is Percentage Change in Pirce. Estimated by OLS with year dummies.
95% Confidence Interval for the coefficients are presented in the Table.

5.3 Structural Difference Across Region?

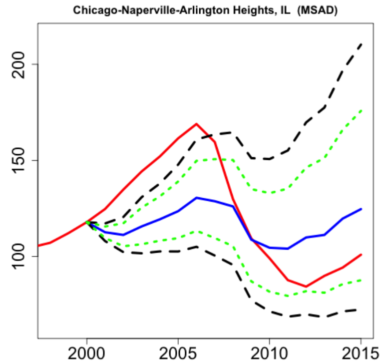
One might worry about that the fundamentals might impact the housing price differently in the four regions in US. In order to test it, I will allow each independent variables in equation (7) to have different coefficients for different regions. I need each fundamental variables to become a interaction term with the regions (One variable becomes four variables in the model). I will also include the year dummies into equation (7) and get estimation by OLS. Table 5 presents the 95% confidence intervals for the coefficients that I allow to be varied by regions. The 95% confidence intervals for the coefficients of percentage change in real rent do not have a overlapping region, and the same thing can be found for change in unemployment rate, and the percentage change in real structure cost. However, given that I have a small sample of 50 MSAs/MSADs, and only 9 MSAs/MSADs are in the northeast region, so the coefficient estimate might not be precise. Since the difference are so dramatic, it is justifiable to pool them together and hold the restriction that the fundamentals do not impact the house market in the four regions differently.

6 Where are the Bubbles?

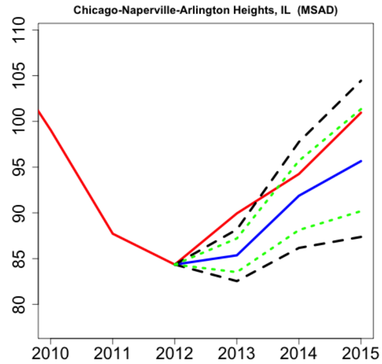
In the literature review, I have presented the definition of a bubble in the housing market: when the price is high enough that the fundamental cannot seem to justify it, then a bubble exists. In another word, if the price in the local housing market is systematically higher than what the fundamentals predict in a extreme way, the existence of bubble can be justified. From the year random effect estimator in Table 4, I can calculate the expected growth rate for each MSA/MSAD in all the years in my sample. Moreover, Stata also gives me the standard error for the fitted value without the time random effect, which I can use to construct the 80% and 95% confidence interval for the expected growth rate that is caused by the changes in fundamentals (the fitted value). Having the confidence interval allows the expected growth rate to include the random market fluctuation into consideration. With the expected growth rates and their 80% and 95%



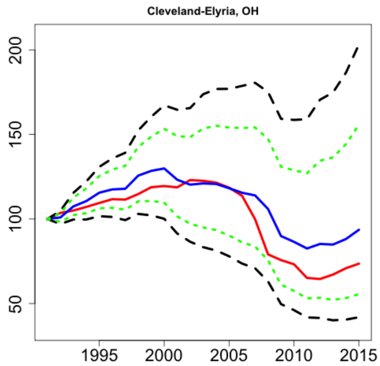
(a) Chicago91



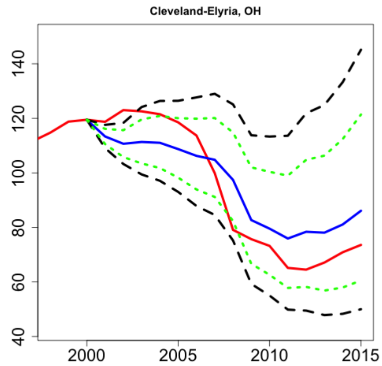
(b) Chicago00



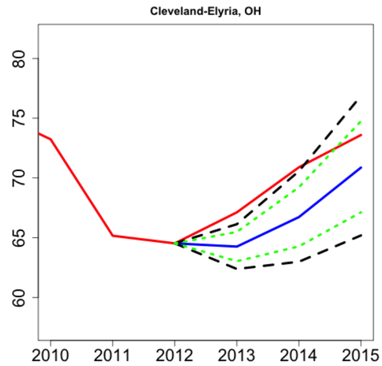
(c) Chicago12



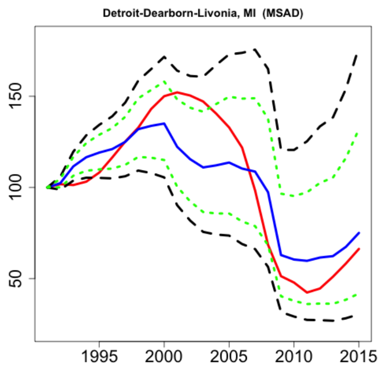
(d) Cleveland91



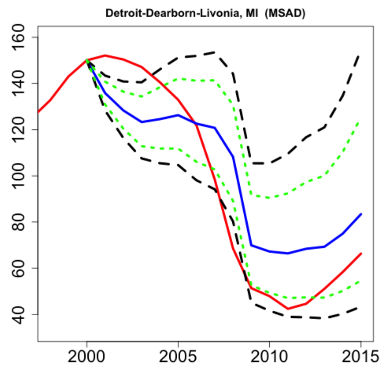
(e) Cleveland00



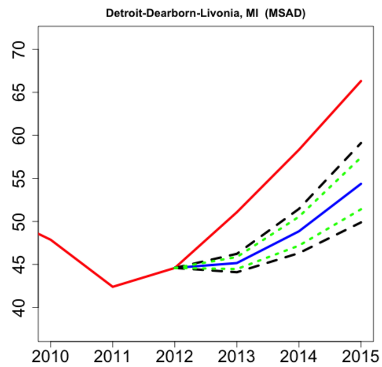
(f) Cleveland12



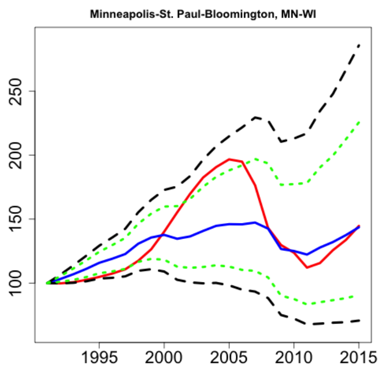
(g) Detroit91



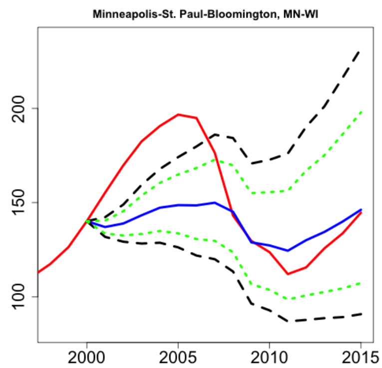
(h) Detroit00



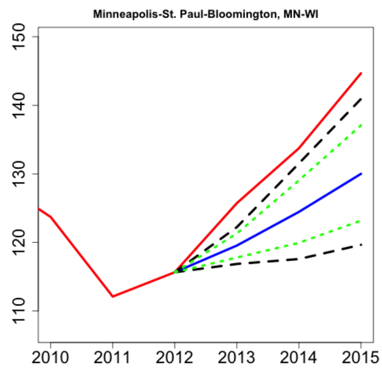
(i) Detroit12



(j) Minneapolis91

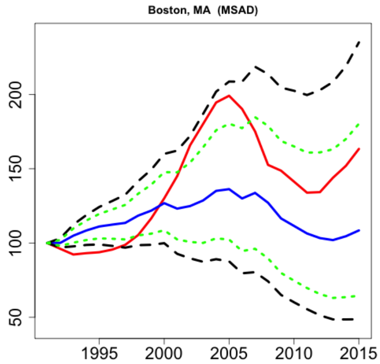


(k) Minneapolis00

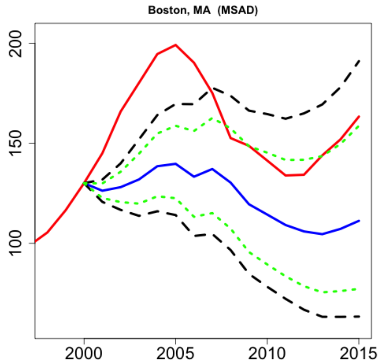


(l) Minneapolis12

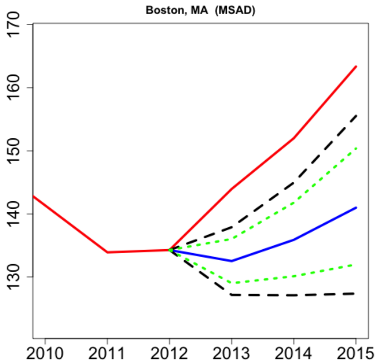
Figure 6: Selected MSAs/MSADs in the Midwest



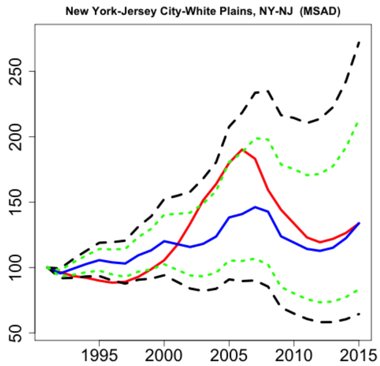
(a) Boston91



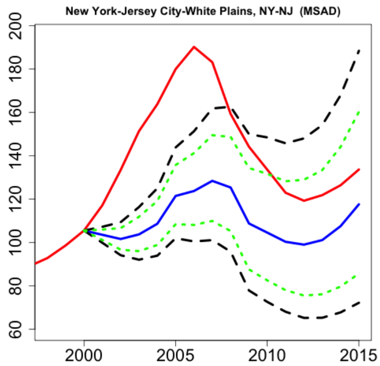
(b) Boston00



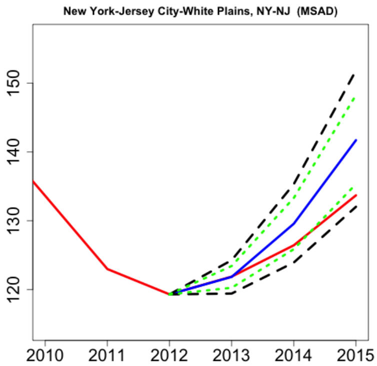
(c) Boston12



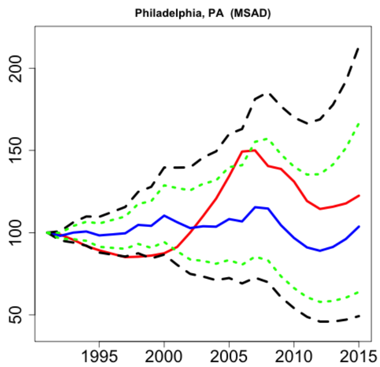
(d) New York - Jersey City91



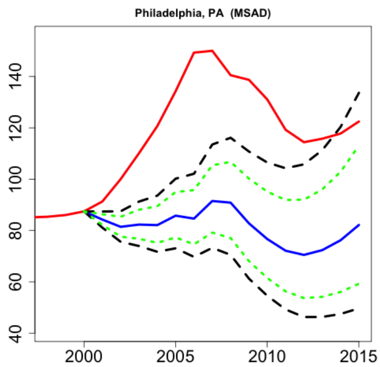
(e) New York - Jersey City00



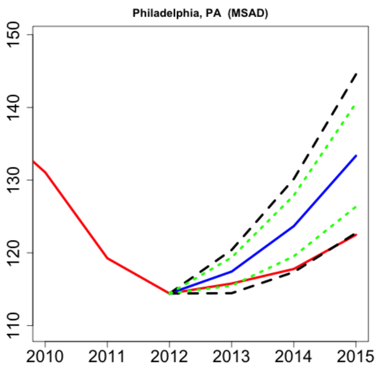
(f) New York - Jersey City12



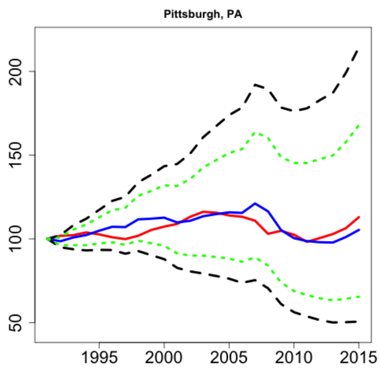
(g) Philadelphia91



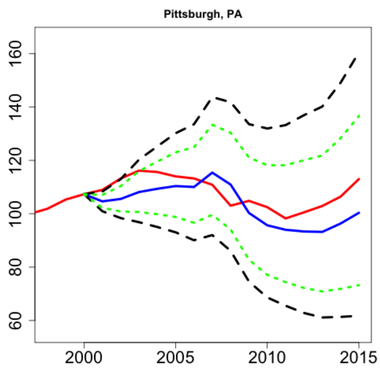
(h) Philadelphia00



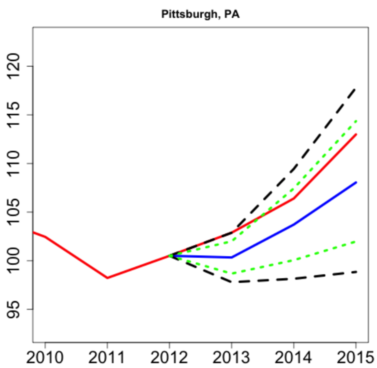
(i) Philadelphia12



(j) Pittsburgh91

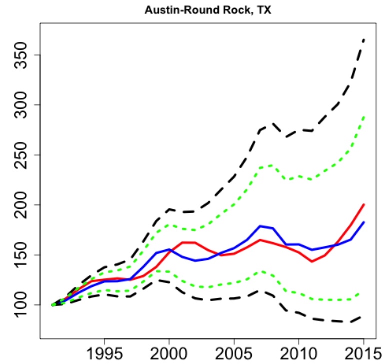


(k) Pittsburgh00

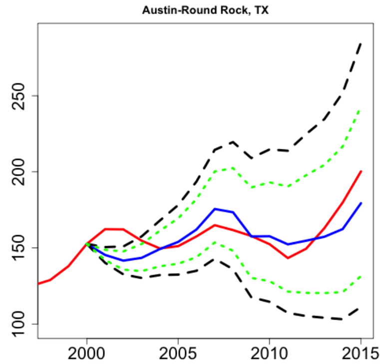


(l) Pittsburgh12

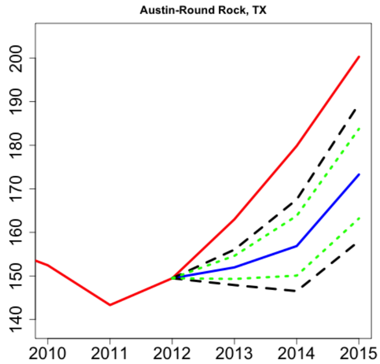
Figure 7: Selected MSAs/MSADs in the Northeast



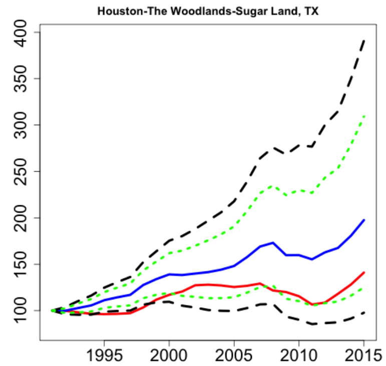
(a) Austin91



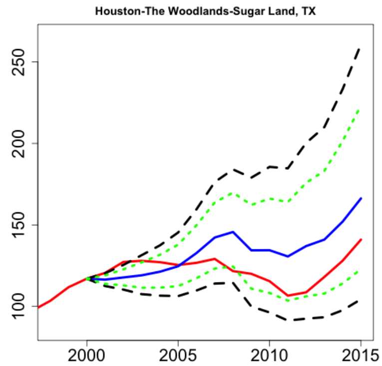
(b) Austin00



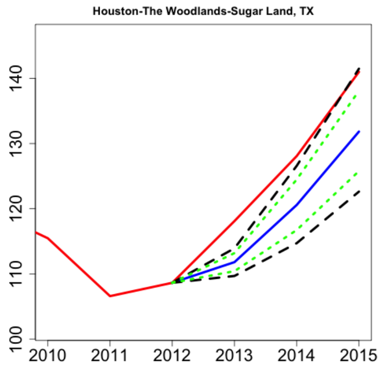
(c) Austin12



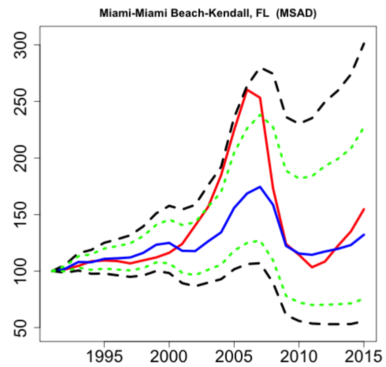
(d) Houston91



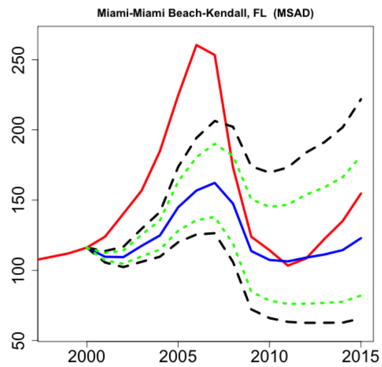
(e) Houston00



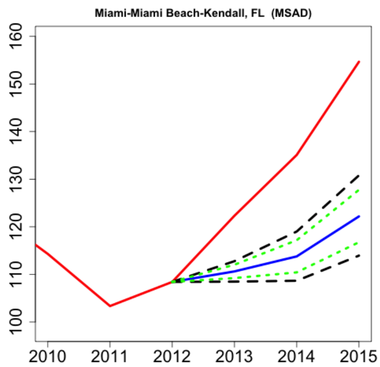
(f) Houston12



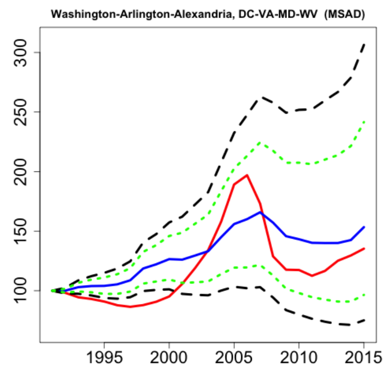
(g) Miami91



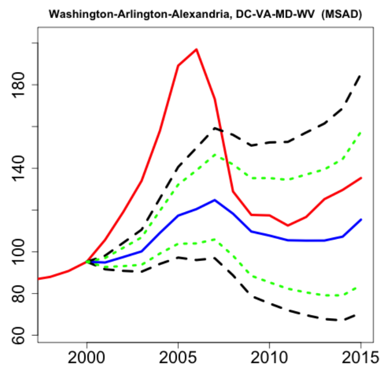
(h) Miami00



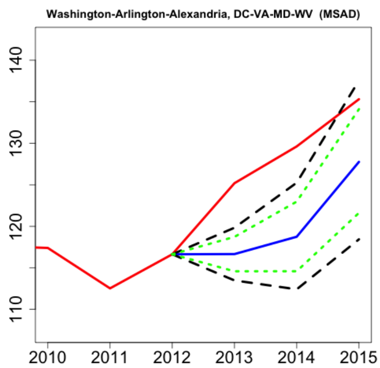
(i) Miami12



(j) Washington91

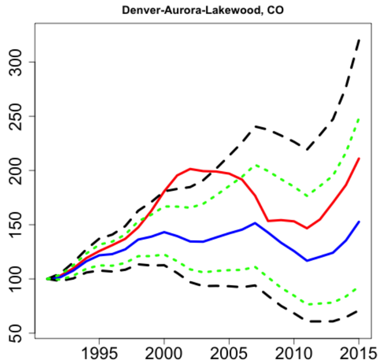


(k) Washington00

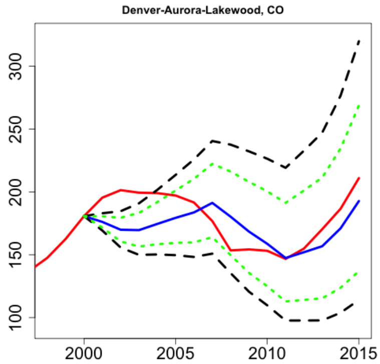


(l) Washington12

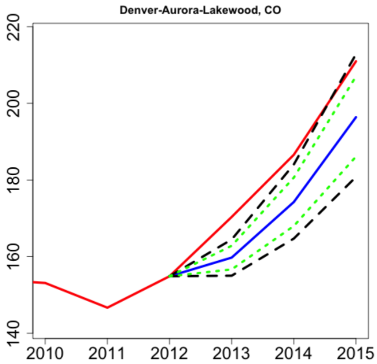
Figure 8: Selected MSAs/MSADs in the South



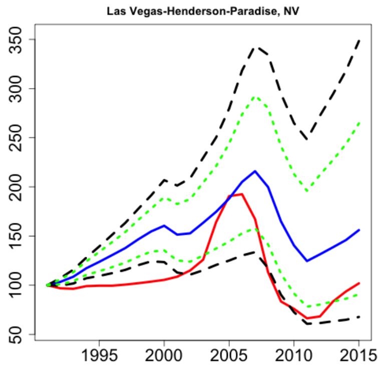
(a) Denver91



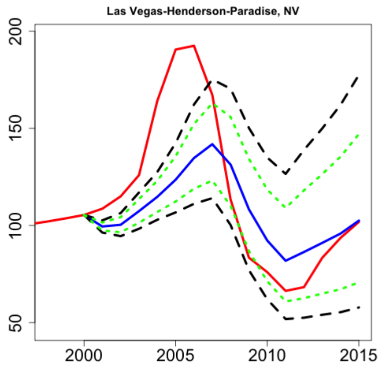
(b) Denver00



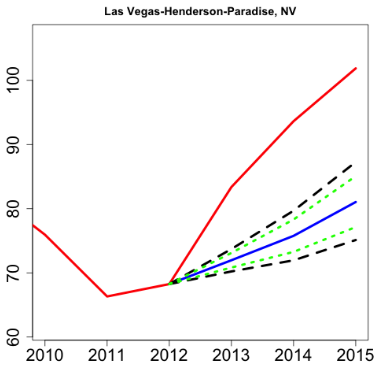
(c) Denver12



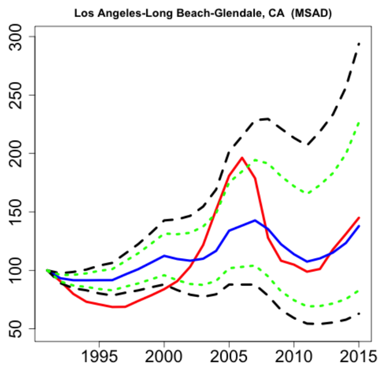
(d) Las Vegas91



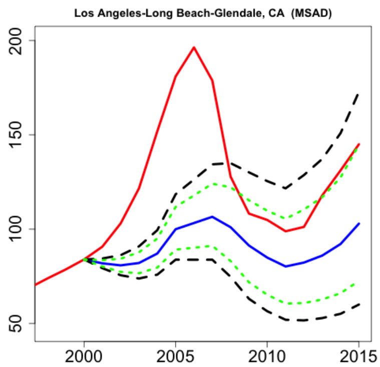
(e) Las Vegas00



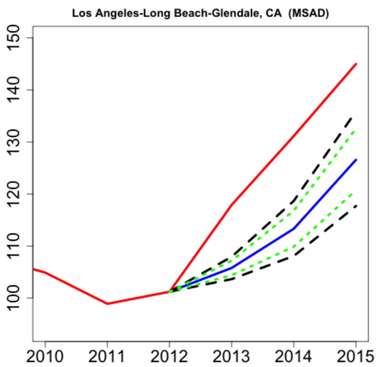
(f) Las Vegas12



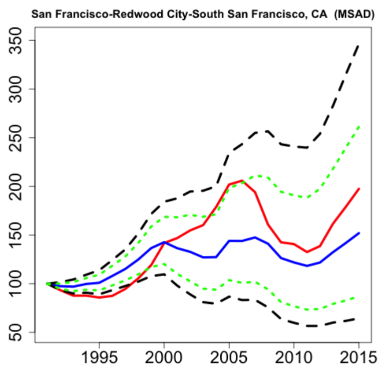
(g) Los Angeles91



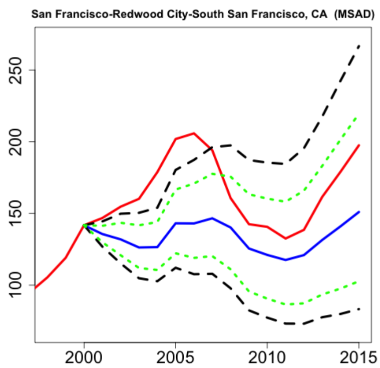
(h) Los Angeles00



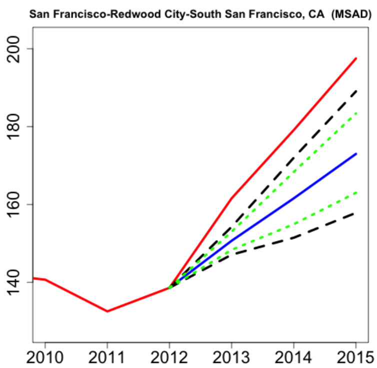
(i) Los Angeles12



(j) San Francisco91



(k) San Francisco00



(l) San Francisco12

Figure 9: Selected MSAs/MSADs in the West

confidence interval, I can construct a the expected house price index series for each MSA/MSADs (blue line in Figure 6, 7, 8, 9) and its 80% (green dotted line) and 95% (black dotted line) confidence band [19]. If the actual real house price index (red line) for a MSA/MSAD systematically lies outside the 95% (or 80%) confidence band, we might suspect the existence of bubbles in the that MSA/MSAD's housing market.

Figure 6, 7, 8, and 9 presents the comparisons for selected MSAs/MSADs in the four regions. The left hand side column in each figure has a reference point of 1991, meaning that I start the calculation for the expected house price index and its confidence bands from year 1991. However, due to the construction method I use for the confidence bands, starting at 1991 does not reveal much information about whether there were bubbles even in the boom from 2000 to 2006 (but clearly there were in many MSAs in retrospect). From the 16 MSAs/MSADs I selected to present, only Denver's real house price index lies outside of the upper 95% confidence band for several years (Figure 9a). Moreover, there were another less severed housing burst due to the saving and loans crisis at 1989, so there was a declining trend in the early 90s' market. Using 1991 as a reference point will certainly underestimate the potential of bubble in the early 2000s. Goodman and Thibodeau (2008) [5] has claimed that the period between 1990 to 2000 is free of bubbles, so I choose 2000 as another reference point which are presented in the middle column of each figure. By using 2000 as a reference point, we can see strong indication of bubbles in the housing market for Chicago, Minneapolis, Boston, New York-Jersey City, Philadelphia, Miami, Las Vegas, Los Angeles, and San Fransisco given that their real house price index is always above the upper 95% confidence band over the period from 2000 to 2006. The comparatively sharper decline in the housing prices for the "bubble" MSAs/MSADs starting in 2007 justify my claim. Moreover, 32 out of the 50 MSAs/MSADs had bubbles in the housing boom from 2000 to 2006 and most of them are in costal regions. This is in consensus with the findings in many literature: the bubbles are mostly concentrated in the coastal region.

To test for the bubble in the recent housing market, we need a new reference point since the expanded nature of the confidence bands will underestimate the potential of bubble in the recent market. The new reference point I choose is 2012, and the plots are presented in the right-hand-side column in each figure. However, given that the housing market experienced a steepest decline not too long ago, it is potentially the house prices are undervalued in some markets in 2012. The strong appreciation could be reflecting that the market is returning to its expected level, and we might overestimate the existence of the bubble in some market, e.g. Detroit (Figure 6i). In order to handle the above issues and being conservative in testing for the bubble, I use the expected house price index (blue line) that use reference year 2000 as indication for whether house price was undervalued in 2012. If the real house price (red line) is below the expected house price (blue line) in 2012 (middle column), then I consider the house price was undervalued in that market. For the markets that were not undervalued in 2012 but show significant appreciation (real house price index consistently lies above the upper 95% confidence interval of the expected house price index from 2012 to 2015), there are potentially bubbles. The MSAs/MSADs that are claimed to have a bubble by the above justification in the 16 selected MSAs/MSADs are Boston, Austin, Miami, Los Angeles, and San Francisco (Denver is not one of them because in 2015, the real house price index is slightly below the upper 95% confidence band). Over the 50 MSAs/MSADs in my sample, only 12 of them have bubbles in the recent housing markets, and all of them are coastal MSA/MSAD except for Austin, TX.

However, people might potentially change their perspective about where the housing price level should be over the year of housing bursting. In the Las Vegas example (Figure 9d), if people do think the average house price level should be consistency over the sample period, then the house price in Las Vegas should always be considered undervalued. We cannot even talk about a bubble in the Las Vegas's market, but there was actually one. 2012 is 2 years away from the Great Recession. If the housing price was undervalued, the readjustment should have taken place over the 2-year-period. For being liberal in testing for bubbles, we can think of the housing price in 2012 as what people consider the normal price is. Then if the real house price (red line) is always above the upper 95% confidence band (top black line) from 2012 to 2015, I will consider that there is a bubble in that market. For the 50 MSAs/MSADs in my sample, 23 of them are considered to have a bubble in this liberal approach (most are MSAs/MSADs in the South and West). The full list of cities is available in Table 6.

7 Conclusion

Using a panel of 50 MSAs/MSADs with data from 1991 to 2015, I examine how the local economic fundamentals - population, real household income, real rent price, unemployment rate, effective mortgage rate, real structure cost, real land value, and price uncertainty - affect the real housing price in the local markets by using a year random effect model. Moreover, by including the regulatory and geographic constraints with allowing different coefficients in boom, burst, and other period into the year random effect model, I find that residential land regulation and geographic land scarcity were related to amplitude of the housing boom and burst in the first decade, strongly supporting the conclusion in Huang and Tang's study (2012) [8].

More importantly, by using the year random effect estimator for the economic fundamentals, I can estimate how much of the annual percentage change in real house price is attributed to the economic fundamentals and how much can be attributed to the unobserved shock (For example, "bubble" in the boom period). By comparing the real house

Table 6: Where are the Bubbles?

2000 - 2006	2012 - 2015 Conservative	2012 - 2015 Liberal
Midwest		
Minneapolis-St. Paul-Bloomington, MN-WI		Minneapolis-St. Paul-Bloomington, MN-WI
Milwaukee-Waukesha-West Allis, WI		Detroit-Dearborn-Livonia, MI
St. Louis, MO-IL		Warren-Troy-Farmington Hills, MI
Northeast		
Boston, MA	Boston, MA	Boston, MA
Cambridge-Newton-Framingham, MA	Cambridge-Newton-Framingham, MA	Cambridge-Newton-Framingham, MA
Montgomery County-Bucks County-Chester County, PA		
Nassau County-Suffolk County, NY		
New York-Jersey City-White Plains, NY-NJ		
Newark, NJ-PA		
Philadelphia, PA		
Providence-Warwick, RI-MA		
South		
Fort Lauderdale-Pompano Beach-Deerfield Beach, FL	Fort Lauderdale-Pompano Beach-Deerfield Beach, FL	Fort Lauderdale-Pompano Beach-Deerfield Beach, FL
Miami-Miami Beach-Kendall, FL	Miami-Miami Beach-Kendall, FL	Miami-Miami Beach-Kendall, FL
	Austin-Round Rock, TX	Austin-Round Rock, TX
Orlando-Kissimmee-Sanford, FL		Orlando-Kissimmee-Sanford, FL
Tampa-St. Petersburg-Clearwater, FL		Tampa-St. Petersburg-Clearwater, FL
West Palm Beach-Boca Raton-Delray Beach, FL		West Palm Beach-Boca Raton-Delray Beach, FL
Baltimore-Columbia-Towson, MD		Atlanta-Sandy Springs-Roswell, GA
Jacksonville, FL		Dallas-Plano-Irving, TX
Virginia Beach-Norfolk-Newport News, VA-NC		
Washington-Arlington-Alexandria, DC-VA-MD-WV		
West		
Anaheim-Santa Ana-Irvine, CA	Anaheim-Santa Ana-Irvine, CA	Anaheim-Santa Ana-Irvine, CA
Los Angeles-Long Beach-Glendale, CA	Los Angeles-Long Beach-Glendale, CA	Los Angeles-Long Beach-Glendale, CA
Oakland-Hayward-Berkeley, CA	Oakland-Hayward-Berkeley, CA	Oakland-Hayward-Berkeley, CA
Riverside-San Bernardino-Ontario, CA	Riverside-San Bernardino-Ontario, CA	Riverside-San Bernardino-Ontario, CA
Sacramento–Roseville–Arden-Arcade, CA	Sacramento–Roseville–Arden-Arcade, CA	Sacramento–Roseville–Arden-Arcade, CA
San Francisco-Redwood City-South San Francisco, CA	San Francisco-Redwood City-South San Francisco, CA	San Francisco-Redwood City-South San Francisco, CA
Seattle-Bellevue-Everett, WA	Seattle-Bellevue-Everett, WA	Seattle-Bellevue-Everett, WA
Las Vegas-Henderson-Paradise, NV		Las Vegas-Henderson-Paradise, NV
Portland-Vancouver-Hillsboro, OR-WA		Portland-Vancouver-Hillsboro, OR-WA
San Diego-Carlsbad, CA		Phoenix-Mesa-Scottsdale, AZ
San Jose-Sunnyvale-Santa Clara, CA		

price index to the 95% confidence band of the expected house price index at reference year of 2000, I find evidence of housing bubble in 32 MSAs/MSADs in my sample, and most of them are in coastal region. This is in consensus with the findings in Goodman and Thibodeau's study (2008) [5] and many other literature. My findings expand the body of related literature by investigating into the period after 2012. Using the same method in testing for bubble in the previous boom at reference year of 2012, I find evidence of housing bubbles in 23 MSAs/MSADs in my sample and they are concentrated in the South and West. Moreover, by considering that the house price might be undervalued in 2012 for some markets given that the recession was not over too long ago, I use a more conservative approach to test for bubble. I still find evidence of bubbles in 12 MSAs/MSADs in my sample and 7 of them are in the West. Austin is the only city that was shown to not have bubble in the housing boom from 2000 to 2006 but present strong evidence of bubble in the recent market in the conservative approach.

By investigating data previous to 2003, Case and Shiller (2003) [2], Himmelberg, Mayer and Sinai (2005) [7], McCarthy and Peach (2004) [10] did not find evidence of bubbles in the housing boom since 2000. If we consider the new housing boom we are observing currently starts in around 2012, including only 3 years of boom period data in my sample and finding supportive evidence of bubble in some MSAs/MSADs conveys an very important message to the housing market. An analysis by the Economist claimed that since the Federal Law is silent on loan-to-value limits for borrowers, one fifth of all loans granted since 2012 having LTV ratios of 95%. If the house prices fall by 5%, then those homeowners will be underwater [20]. A 5% decline in house price will easily occur if the bubbles become unsustainable in the markets. Since the previous housing burst has caused catastrophic impact on both the housing market and national economic, national and local policy makers have to take actions to avoid another crash given that the sign of bubble has been seen in some of the local markets. More concerns are needed for markets with stricter regulation and less developable land.

Notes

[1] https://www.bettermarkets.com/sites/default/files/Better\%20Markets\%20-\%20Cost\%20of\%20the\%20Crisis.pdf

[2] http://www.economist.com/news/leaders/21705317-americas-housing-system-was-centre-last-crisis-it-has-still-not-been-properly

[3] I decide to use annual building permit data rather than monthly because building permit is highly seasonal.

[4] http://www.fhfa.gov/datatools/downloads/documents/hpi_focus_pieces/2011q2_hpifocus_n508.pdf

[5] Population data is available from 1990 to 2015. Median household Income is available from 1989 to 2015 except for 1990, 1991, 1992, 1994, 1996. I use a smooth spline fit available on R to estimate the median household income at the missing year.

[6] https://www.fhfa.gov/Media/PublicAffairs/Pages/Housing-Price-Index-Frequently-Asked-Questions.aspx

[7] Suppose we do not have data on MSA/MSAD i and the metropolitan areas A and B are within 300 miles to MSA/MSAD i. The distance between A and i is d_A and the distance between B and i is d_B. Then the weight is $\frac{1}{d_A}/(\frac{1}{d_A}+\frac{1}{d_B})$ for metropolitan A's mortgage rate, and the weight is $\frac{1}{d_B}/(\frac{1}{d_A}+\frac{1}{d_B})$ for metropolitan B's mortgage rate.

[8] http://www.usda.gov/nass/PUBS/TODAYRPT/land0815.pdf

[9] The data set also provide estimated land value for residential development. The land value is calculated by the estimated home value minus the estimated structure cost. If I include both structure cost and land value into my data set, I may run into problem when I run regression. It is not perfect collinearity issue since I do not use their estimated home value as dependent variable, but potential over-fitting problem will exist.

[10] The Wharton Regulation Index in Saiz's paper does not match perfectly with the index in Gyourko, Saiz, and Summers' paper [6], but I will go with Saiz's data in my study.

[11] Sacramento is the only MSA that I cannot find a match in Saiz's paper, so I assign the index of Stockton-Lodi, CA (Sacramento's nearby MSA) to Sacramento. I check on Google map that the two MSA look similarly in geology. The two nearby MSA is possible to share similar regulation rule.

[12] For unemployment rate, I first construct the annual unemployment population and labor force population from the monthly data and then divide the annual unemployment population by labor force and multiple by 100 to get the annual unemployment rate.

[13] $(1+47.76\%)/(1+25.71\%)-1=17.3\%$.

[14] Detroit MSA includes Detroit-Dearborn-Livonia MSAD and Warren-Troy-Farmington Hills MSAD in my sample. Both experience deprecations in the boom and burst period but a strong appreciation in the recent market. It is highly possible that the house values for the two markets are undervalued after the burst.

[15] The Wharton Regulation Index is designed to have mean 0 and standard deviation of 1 across the US cities.

[16] Hwang and Quigley [9] do not take the difference for the uncertainty in price, and I decide to follow them by not using the difference.

[17] I cannot find a good estimation for expected gain given that I include the boom and burst period in my sample. Some literature suggest using moving average of the growth rate, but including it might over-fit the model because of the boom and burst period in my sample.

[18] The estimated χ^2 statistics is 7111.76 which corresponds to a p-value of 0.000 at 1 degree of freedom, so I can reject the null hypothesis. I also conduct a Hausman test. The χ^2 statistics is 12.21 which

corresponds to a p-value of 0.1419 at 8 degrees of freedom, suggesting a random effect model is in favor as compared to a fixed effect model.

[19] For example, if the estimated growth rate for MSA i in year t is a, to calculate the expected house price index at year t is by the following equation: $P_{i,t} = (1 + a/100)P_{i,t-1}$. To construct the upper 95% confidence band, I replace a by the upper 95% confidence interval for the estimated growth rate in each time period in the above equation when doing the calculation. Same apply for the other confidence band calculation.

[20] http://www.economist.com/news/briefing/21705316-how-america-accidentally-nationalised-its-mortgage-market-comradely-capitalism

References

[1] Radha Bhattacharya and Sei-Wan Kim. "Economic fundamentals, subprime lending and housing prices: Evidence from MSA-level panel data". In: *Housing Studies* 26.6 (2011), pp. 897–910.

[2] K.E. Case and R.J. Shiller. "Is there a bubble in the housing market? (with comments and discussion)". In: *Brookings Papers on Economic Activity* 2 (2003), pp. 299–342.

[3] Steven P Clark and T Daniel Coggin. "Was there a US house price bubble? An econometric analysis using national and regional panel data". In: *The Quarterly Review of Economics and Finance* 51.2 (2011), pp. 189–200.

[4] Morris A Davis and Michael G Palumbo. "The price of residential land in large US cities". In: *Journal of Urban Economics* 63.1 (2008), pp. 352–384.

[5] Allen C Goodman and Thomas G Thibodeau. "Where are the speculative bubbles in US housing markets?" In: *Journal of Housing Economics* 17.2 (2008), pp. 117–137.

[6] Joseph Gyourko, Albert Saiz, and Anita Summers. "A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index". In: *Urban Studies* 45.3 (2008), pp. 693–729.

[7] Charles Himmelberg, Christopher Mayer, and Todd Sinai. "Assessing high house prices: Bubbles, fundamentals and misperceptions". In: *The Journal of Economic Perspectives* 19.4 (2005), pp. 67–92.

[8] Haifang Huang and Yao Tang. "Residential land use regulation and the US housing price cycle between 2000 and 2009". In: *Journal of Urban Economics* 71.1 (2012), pp. 93–99.

[9] Min Hwang and John M Quigley. "Economic fundamentals in local housing markets: evidence from US metropolitan regions". In: *Journal of Regional Science* 46.3 (2006), pp. 425–453.

[10] Jonathan McCarthy and Richard W Peach. “Are home prices the next bubble?” In: *Economic Policy Review* 10.3 (2004).

[11] Vyacheslav Mikhed and Petr Zemčık. “Do house prices reflect fundamentals? Aggregate and panel data evidence”. In: *Journal of Housing Economics* 18.2 (2009), pp. 140–149.

[12] Albert Saiz. “The geographic determinants of housing supply”. In: *quarterly Journal of Economics* 125.3 (2010).

[13] Joseph E Stiglitz. “Symposium on bubbles”. In: *The Journal of Economic Perspectives* 4.2 (1990), pp. 13–18.

[14] Wei-Xing Zhou and Didier Sornette. “Is there a real-estate bubble in the US?” In: *Physica A: Statistical Mechanics and its Applications* 361.1 (2006), pp. 297–308.

Appendix A Limitation and Further Research

First, I use the percentage change in real farm estate value as a proxy for land value in the model. Given that many MSAs/MSADs are not heavily agricultural (e.g. New York City), if people do not use the land for residential purpose, they might use it as a parking lot or other infrastructures rather than using it for agriculture. Another issue is that there is not many farm lands in a MSA and the farm real estate value I got is on state level. There is certainly some measurement error of using it as a proxy. One obvious example is the Nassau County-Suffolk County MSAD in New York State. The New York State itself is full of agricultural land so the value is cheap. However, Nassau County-Suffolk County (Long Island) certainly have higher land value than the vast New York state does. Given the limited time I have, this is the best one I could find so far. One future research direction is to find a better proxy for change in land value in MSA level.

Second, the data I have is a panel time series data. However, when I do the modeling, I did not take much advantage of treating it as time series data because of my limited exposure to the panel time series technique. The future direction is to learn more about panel time series data and hopefully it can help me come up with a better model.

Third, to collect data on county or census tract level so I can include a MSA year interaction into the model. Then the coefficient itself can tell whether a bubble exists. However, census tract information is not updated annually and it is very difficult to get census tract data dated back to the 90s.

Last, when I interpret the bubble, I can only say there is a bubble or there is not in a given time period. However, in reality, when we think about bubble, we think about the probability of having a bubble in the market. Bayesian statistics can help when dealing with probability. The bubble can be modeled as a latent variable that follows a Bernoulli distribution in each time period. If there is a bubble in period t, we are likely to observe either a bubble or a burst in period $t + 1$. By using a Hidden Markov Model, we can not only predict the probability of whether a bubble exist in period t, but also predict the potential turning point for the bubble. However, math is needed to work out the Gibbs Sampling before training the data.