When Real Estate is the Only Game in Town*

Hyun-Soo Choi† Harrison Hong‡ Jeffrey D. Kubik§

Jeffrey P. Thompson¶

First Draft: January 15, 2013

This Draft: Dec 24, 2015

Abstract

The excesses of the historic US housing cycle of the 2000s were concentrated in the Metropolitan Statistical Areas (MSAs) of Nevada, Phoenix, Florida, and California. Even controlling for factors associated with leading explanations of this housing cycle such as differences across MSAs in the amount of subprime lending and their housing supply elasticities, these Sand State MSAs had more than double the mortgage origination, price appreciation and subsequent crash as other MSAs. We show that these excesses, can however, be explained by Sand State MSAs having an abnormally low supply of public equities (measured using the book value of firms headquartered there) relative to total income. Because households are locally biased in their investments, they purchase investment homes nearby rather than owning stocks, amplifying the housing cycle in these MSAs.

*Choi acknowledges support from the Sim Kee Boon Institute, Singapore Management University. We thank Tim Loughran, Chris Mayer, Wenlan Qian, Jeremy Stein, Nada Mora, and seminar participants at the EFA 2015 annual conference, UC Berkeley, St Gallen, Dartmouth, Notre Dame, National University of Singapore, the CICF 2013 conference, the NBER Behavioral Finance meeting, KAIST, Yonsei, SKKU, the ABFER Annual conference 2014 for helpful comments. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff of the Federal Reserve or the Board of Governors.

†Singapore Management University and SKBI (e-mail: hschoi@smu.edu.sg)

‡Princeton University, NBER and CAFR (e-mail: hhong@princeton.edu)

§Syracuse University (e-mail: jdkubik@maxwell.syr.edu)

¶Federal Reserve Board (e-mail: jeffrey.p.thompson@frb.gov)

1. Introduction

A defining feature of the US housing cycle of the 2000s, which contributed greatly to the Great Financial Crisis and Recession, is that its excesses were concentrated in the states of Arizona, Nevada, Florida and California. Because these states are not contiguously located but share an abundance of beaches or deserts, housing market analysts have labeled them Sand States. MSAs in Sand States had more than double the mortgage originations over the period of 2000-2006 and significantly larger home price appreciation and subsequent crash as any other MSA (see, e.g.,Olesiuk and Kalser (2009)).

Three main explanations have been widely discussed regarding this recent US housing cycle. The first is ease of lending standards for low income or subprime households, especially those living in low income growth areas (Mian and Sufi (2009), Keys et al. (2009), Palmer (2014)). The second is housing supply inelasticity, whereby the difficulty of building homes due to land or zoning issues might have amplified the housing price volatility (Glaeser et al. (2008)).[1] The third is irrational exuberance, as a rise in home prices perhaps initially brought on by fundamentals gets over-extrapolated (Shiller (2005), Glaeser et al. (2013)).

But none of these can easily account for the Sand States. For instance, we find that low income growth MSAs in the 1990s and low housing supply elasticity MSAs had bigger housing cycle excesses, consistent with the importance of subprime lending and difficulty of building homes. But in a multiple regression, they do not explain much of the Sand States effect. Similarly, other fundamental factors such as population growth in certain MSAs in the 1990s, which might lead to over-extrapolation, also cannot rationalize the excesses of the Sand States.

Our perspective is that it is difficult to get a full grip of the causes of the historic US housing cycle of the 2000s without understanding what drove these excesses. A key fact which motivates our analysis is that households' investments, be it stocks or non-owner

[1] The rationale is that in low housing supply elasticity MSAs, supply could not quickly adjust to rising demand and so prices had to. In contrast, in high housing supply elasticity MSAs, price responses were more muted as many more homes were built.

occupied homes, are locally biased. Driven by familiarity or other informational frictions, households do not diversify but rather hold concentrated positions in stocks headquartered within 60 miles of where they live (see, e.g., French and Poterba (1991), Grinblatt and Keloharju (2001), Huberman (2001)). This local bias is even more stark when it comes to non-owner occupied homes bought for investment.

From an annual National Association of Realtors (NAR) national survey of home purchases, households characterize these purchases in the same way that they describe their purchases of stocks. Investment homes are bought to generate rental income and typically face higher interest rates and more stringent collateral requirements than either primary residences or vacation homes because banks view them as speculative investments akin to stocks. Panel A of Table 1 reports the distribution of the distances of the second home purchases from the primary residence of the buyer. For instance, the median distance of the investment home from the buyers' primary residence in 2005 was 10 miles. In contrast, the median distance of vacation homes was 220 miles.[2] These numbers are very stable across the many years over which the surveys were conducted.

As a result of this local bias, we propose an explanation having to do with the local investment opportunity sets of investors. Our analysis is particularly interested in the supply of publicly available firms in an MSA, measured using the book equity of firms headquartered in an MSA (BOOK VALUE) relative to the total income of that MSA (a RESIDUAL BOOK VALUE variable which we construct from a regression of BOOK VALUE on total income across MSAs). Sand State MSAs have on average low RESIDUAL BOOK VALUE, meaning a low supply of equities relative to their income. These values are measured in the 1990s as well, similar to income growth and population growth, because we want to explain future mortgage originations and price movements comparing these variables.

[2]Note that a strong local bias of investment homes is not inconsistent with the well-known findings in Chinco and Mayer (2012) that out-of-town buyers drove real estate markets like Phoenix and Las Vegas. Out-of-town buyers need not be buying investment homes and might be moving to these cities instead. Their study examines price implications and do not have household portfolio data. The fractions of out-of-town buyers are in any event typically a small fraction of total purchases.

We hypothesize that as a result households in MSAs with few local stocks headquartered there turn to nearby investment homes as a substitute. We call this an only-game-in-town effect. Because households in low RESIDUAL BOOK VALUE MSAs have a higher underlying propensity to buy investment homes, we expect that any given national factor, be it easy lending standards or aggregate optimism, gets magnified in these MSAs as the households there ought to buy a greater number of investment homes. And mortgage originations and home prices ought then to be more cyclical as a result.

The role of investment homes in driving the Sand States phenomenon is an ex-ante plausible one. According to the same National Association of Realtors (NAR) Home Buyers Survey, investment homes represented on average about 22% of the residential sales market nationally or around one million homes annually between 2003 and 2013. Beyond being a sizeable part of the real estate market, and as can be seen in Panel B of Table 1, demand for these homes seems to be correlated with the recent housing cycle of 2002-2007. In 2003, primary residences accounted for 67% of the market, vacation properties 12% and investment homes 22%. During the peak year of 2005, investment homes rose to 28% of the market while primary residences dropped to 60%.

Indeed, we show that the RESIDUAL BOOK VALUE variable can account for a significant fraction of the excesses of the Sand States in multiple regressions through this investment home channel. First, it can explain around 50% of the total mortgage originations in the Sand States. Second, and consistent with our hypothesis, this is largely due to the ability of RESIDUAL BOOK VALUE to pick up mortgage originations of investment homes. Third, it can account for about 20% of the home price fluctuations, both the rise and subsequent crash. These regressions involve all 277 MSAs and are population weighted. Nonetheless, a large fraction of these MSAs have population around 100 thousand people, while large MSAs might arguably be of more interest. When we focus just on large MSAs with a population greater than or equal to 750 thousand people, which is 65 MSAs in total, we get even stronger results. Our RESIDUAL BOOK VALUE variable can explain virtually all of the

Sand States effect.

Our measure generalizes Hong, Kubik, and Stein (2008)'s RATIO of BOOK VALUE to total income at the state level. They show that RATIO is inversely related to the price-to-book of companies in that state. That is, states with fewer companies have higher prices for their stocks due to an only-game-in-town effect. In the setting of large MSAs, we can use the simpler RATIO measure and find similar results as our RESIDUAL BOOK VALUE measure. The reason is the small MSAs have no firms and hence automatically have a RATIO of zero. But this is not an issue for large MSAs which all have some firms headquartered there.

We then bring to bear new data to test the key mechanism behind our hypothesis — namely that households living in low RESIDUAL BOOK VALUE or low RATIO MSAs buy more investment homes. We use data on household investment portfolios from the Federal Reserve Board's Survey of Consumer Finances (SCF). The SCF samples a cross-section of roughly 5,000 to 6,000 households once every few years. The primary strength of the SCF is its sampling of the right tail of the wealth distribution, which is typically missing in most other surveys. We use the 1995, 1998, 2001, 2004, 2007, and 2010 waves. For instance, we can calculate for each SCF household the value of its investment homes and vacation homes as a fraction of its total assets as well as its investments in stocks as a fraction of total assets (% INVESTMENT REAL ESTATE, % VACATION HOMES, and % DIRECTLY-HELD STOCKS). We also know the MSA, zip-code and county where the household lives and also have a host of demographic information to use in our analysis.

The SCF data is most accurate for large MSAs, and so we will focus on this subset of MSAs where our only-game-in-town effects are largest and use the simpler version of our measure, namely RATIO as the key independent variable of interest. To begin with, we consider how the usual household demographic factors affect investment home and stock ownership. First, we find as expected that real estate investment homes and stocks as a fraction of total assets are both rising in household wealth and income. Second, we find that households living in high average FICO (prime) zip codes, all else equal, own less investment homes and more stocks than counterparts in low average FICO (subprime) zip codes. To a lesser degree, white households, all else equal, also own less investment homes and more stocks than non-white households. There are many reasons for this heterogeneity in preferences from the literature on household finance, including investor sophistication, familiarity with stocks as opposed to homes either due to employment, or other social interaction, attention and cultural biases (Barber and Odean (2008), Guiso et al. (2009)).

Controlling for these factors, we find, consistent with our only-game-in-town hypothesis, that households living in MSAs where there are few stocks headquartered own more investment homes and less stocks compared to other households.[3] The mean percent of household assets in investment real estate is around 3% and the mean percent in directly held stocks is 2%. A one standard deviation increase in log RATIO decreases the percentage of assets held by households in investment real estate by about 40 to 52% relative to this mean. It increases the percentage of total assets held in stocks by about 80% of its mean. In other words, it appears that households turn to substitutes in the form of investment homes nearby when real estate is the only game in town.

One of the main concerns of our empirical analysis is that MSAs with few local firms might simply have more low FICO type households or non-white households who prefer homes to stocks. To rule out this alternative explanation, we use a difference-in-differences strategy. Because there are different types of zip codes (low versus high FICO) within an MSA, we can include Year $\times$ MSA Effects in our panel regressions and interact log RATIO with whether the household belongs to a high FICO zip code. That is, we attempt to measure our only-game-in-town effect by looking at the difference in the behaviors of households living in highest FICO (average score 721) versus lowest FICO zip codes (average score 621) within the same MSA. We then expect that in low RATIO MSAs households in high FICO zip codes will not tilt away from investment homes as much compared to households in low FICO zip

[3]More precisely, we take the log(.00001+RATIO) to account for MSAs that have no stocks located nearby. Our baseline sample focuses on large MSAs with populations greater than 750,000 people so instances of zero RATIO are rare. The results using data from all MSAs are similar.

codes because there are few local stocks for them to invest in. That is, in MSAs with few local firms, the behavior of households in low and high FICO zip codes ought to be more similar than in MSAs with many local firms.

We indeed find that the portfolios of households, even high FICO households who on average have a preference for stocks as opposed to investment homes, are tilted toward investment homes in MSAs with few local firms, consistent with our only-game-in-town effect.[4] Our regression estimates suggest that a one standard deviation decline in log RATIO in a MSA is enough to offset the average level difference in the propensity of low FICO households to hold more investment real estate than high FICO households.

Our findings emphasizing the key role of investment homes in the housing cycle is reminiscent of several interesting studies of this recent cycle (see, e.g., Chinco and Mayer (2012), Li and Gao (2013), and Haughwout et al. (2011)). These papers show that there was significant variation in non-owner occupied home purchases across US Metropolitan Statistical Areas (MSAs) and that those MSAs with the highest fraction of investment home transactions experienced the biggest boom and bust patterns. Our paper provides a systematic rationale for why this was the case.

Our paper proceeds as follows. We first examine in Section 2 the relationship between the Sand State excesses and our only-game-in-town variables. Having established the relevance of these factors, we then turn to the SCF data in Section 3 to study the mechanism behind our variables. We conclude in Section 4. Additional robustness checks and discussions are in an Internet Appendix.

2. Sand State Excesses and Local Supply of Equities

We begin by relating the excesses of the Sand State MSAs to the local supply of firms headquartered there.

[4] We can also consider differences in the behaviors of white versus non-white households within the same MSA by interacting log RATIO with an indicator for whether a household is white; the FICO interaction effect is, however, stronger.

2.1. MSA Panel Data

To start, we construct a panel data set of annual observations on the conditions of the housing market in MSAs from 2000 to 2010. From the Federal Housing Finance Agency (FHFA), we collect estimates of the yearly housing price appreciation of homes within a MSA. We gather yearly changes in various measures of mortgage origination amounts within a MSA from data of the Home Mortgage Disclosure Act (HMDA).[5] The MSA supply elasticity calculated by Saiz (2010) is appended to this panel.

We have complete information on 277 MSAs over our time period. Panel A of Table 2 shows the summary statistics of these data. We break up the statistics into two periods: the housing bubble period of 2000 to 2006 and the crash period from 2007 to 2010. Not surprisingly, we find on average that there were substantial increases in housing prices and mortgage origination across most MSAs in the bubble period, and those increases turn around again in most MSAs in the crash period.

The data necessary to measure the relative supply of local stocks in a MSA similar to Hong et al. (2008) are added to this panel. We use information from COMPUSTAT to calculate the total book value of firms headquartered in a MSA and the Bureau of Economic Analysis (BEA) for the total income in a MSA. We also collect measures of changes in the economic conditions of MSAs during the 1990s, before our analysis. From Census sources, we measure the percentage change in population in a MSA between 1990 and 2000. Again using information from the BEA, we calculate the percentage change in total income during the same period.

Panel B of Table 2 shows the summary statistics of a subsample of the 65 largest MSAs in our sample: those with an average population in the sample period of at least 750,000. For much of our analysis, we will examine the sensitivity of our results when we limit our sample to bigger MSAs to make sure our results are not being driven by small cities. The

[5] We concentrate on two measures of mortgage origination. The first is the percentage change in total mortgage origination in a MSA. The second is the percentage change in second home mortgage origination.

summary statistics of the subsample are very similar to the entire data set.

2.2. Measures of Local Stock Supply

In the spirit of Hong et al. (2008), we want to measure the differences across MSAs in the availability of local stocks. Cities with a large total book value of firms normalized by the cities' total income are considered to have a high supply of local stocks; cities with a low total book value relative to total income have a low supply.

For each MSA in our sample, we measure their average total book value of firms and the average total income from 1996 to 2000. We then regress the log average total book of a MSA on its log average total income from 1996 to 2000. The results of this regression are presented in column (1) of Panel A Table 3. Not surprisingly, there is a strong positive relationship between a city's total firm book value and income. We use the residual of this regression as a measure of the relative supply of local firms in a MSA. In other words, MSAs with high local firm book value that cannot be explained by total income are considered high local supply MSAs. We call this a MSA's RESIDUAL BOOK VALUE.

Because we are not certain what the functional form of the regression of total book value on total income should be, we also try two other specifications. We try adding a log Total Income square term to the original specification, and we also add a cube term. The results of those regressions are presented in columns (2) and (3) in Panel A of Table 3. Adding these additional terms to the regression specification does not change the residual much. Panel B of Table 3 shows the correlation matrix of the three residuals. The correlations are very high. In the rest of our analysis, we use the MSA RESIDUAL BOOK VALUE using the regression that includes the square term.[6]

We first show that MSAs in Sand States on average have low RESIDUAL BOOK VALUEs. Figure 1 shows the estimated relationship between log Total Book Value and log Total Income for the 277 MSAs in our sample. Around that fitted line, we add the actual data

[6]Not surprisingly given the high correlation of the different residuals, our results are almost identical when we use the residuals from the other specifications.

points of each of the MSAs. We highlight the MSAs that are in the Sand States. Over 60% of Sand State MSAs are located below the fitted line, indicating that they have a low supply of local stocks compared to other MSAs.[7]

We show the relationship between RESIDUAL BOOK VALUE and whether the MSA is in a Sand State more formally in Table 4. We regress a MSAs RESIDUAL BOOK VALUE on a dummy variable for the MSA being in a Sand State. Column (1) shows the results of this model. The coefficient on the Sand State dummy is negative and statistically different from zero (t-statistic = -4.0). The magnitude of the coefficient suggests that being a MSA located in a Sand State lowers a MSA's RESIDUAL BOOK VALUE by 1.14. Given that the standard deviation in our sample of RESIDUAL BOOK VALUE is 1.0, this means that Sand State MSAs have on average a lower local stock supply that is 1.14/1.4=81% of a standard deviation of RESIDUAL BOOK VALUE.

We present the results in column (2) of a similar regression, except that now we also add Census Division fixed effects to the specification.[8] With these extra controls, the Sand State dummy is identified by comparing the RESIDUAL BOOK VALUEs of MSAs in Sand States to other MSAs in the same Census Division, instead of comparing them to all other MSAs. The coefficient on Sand State dummy decreases in absolute value compared to column (1), but it is still large. Being a MSA in a Sand State on average lowers the RESIDUAL BOOK VALUE of the MSA by about 0.77/1.4 = 55% of a standard deviation of RESIDUAL BOOK VALUE.

In column (3) of Table 4, we present the results of a regression of the rank of a MSAs RESIDUAL BOOK VALUE on a dummy variable for the MSA being located in a Sand State. We sort the 277 MSAs and assign a rank of 1 to the MSA with the highest residual, 2 to the

[7]Table A1 lists the RESIDUAL BOOK VALUEs of our large MSA sample. The MSAs in Sand States are highlighted.

[8]There are nine Census Divisions. They are New England (CT, ME, MA, NH, RI, VT), Middle Atlantic (NJ, NY, PA), South Atlantic (DE, FL, GA, MD, NC, SC, VA, WV), East South Central (AL, KY, MS, TN), East North Central (IL, IN, MI, OH, WI), West South Central (AR, LA, OK, TX), West North Central (IA, KS, MN, MO, NE, ND, SD), Mountain (AZ, CO, ID, MT, NV, NM, UT, WY), and Pacific (AK, CA, HI, OR, WA).

second highest residual MSA, etc. The coefficient on the Sand State dummy is negative and large, suggesting that MSAs located in the Sand States are far down the residual ranking. The magnitude indicates that on average Sand State MSAs are about 64 places lower down the ranking. Adding Census Division fixed effects to the regression specification in column (4) suggests that on average Sand State MSAs are about 52 places lower down the ranking.

2.3. The Relationship between RESIDUAL BOOK VALUE and the Excesses of the MSA Housing Markets

After showing that MSAs located in Sand States on average have low supplies of local firms, we now need to measure the relationship between changes in the housing market in the 2000s and RESIDUAL BOOK VALUE. We start by concentrating on changes over time in mortgage origination across MSAs. We break our sample into two periods: the bubble period (2000-2006) and the crash (2007-2010). We then measure how various measures of mortgage origination across MSAs are related to local firm supply.

Our first measure of mortgage origination is the average percentage change in total origination in a MSA. We regress this variable on the MSAs RESIDUAL BOOK VALUE, the MSAs supply elasticity, a dummy for the MSA having high growth in income during the 1990s, a dummy for the MSA having high growth in population during the 1990s and state effects.[9]

The estimates using the bubble period sample and all MSAs are presented in column (1), Panel A of Table 5. The coefficient on the RESIDUAL BOOK VALUE is negative and statistically different than zero (t-statistic = -3.3), indicating that MSAs with a smaller supply of local firms experienced a bigger increase in total mortgage origination during the bubble period. The magnitude of the coefficient suggests that a one standard deviation decrease in RESIDUAL BOOK VALUE (1.40) is associated with an increase in the annual

[9] Because some MSAs straddle more than one state, we turn on a state effect for all of the states a MSA is located in. We have also run the models turning on a state effect only in the state the MSA has the most population in. The results of these alternative specifications are very similar to what is presented.

percentage growth of mortgages of -.013×-1.40 = 0.018. Given that the average annual increase in total mortgage origination during this period with 27%, this implies a .018/.27 = 7% increase in the annual growth.

The coefficient on the housing supply elasticity suggests that more elastic MSAs experienced a smaller increase in mortgages during the boom. Also, MSAs that grew more slowly during the 1990s in terms of income experienced a higher growth in mortgages. This is consistent with the literature, which finds that low income growth MSAs in this cycle had more pronounced housing excesses due to subprime lending. This is contrary to what is typical, in that higher income growth areas should have higher price growth. MSAs that had high population growth, however, had larger growth as expected. We will come back to these two factors later on when we see whether they can explain the Sand State effect.

In column (2), we show similar estimates using the crash sample. The coefficient on RESIDUAL BOOK VALUE is now positive and statistically different than zero. MSAs with smaller local firm supply experienced a bigger decline in mortgages than other MSAs during the crash. The size of the coefficient suggests that a one standard deviation decrease in RESIDUAL BOOK VALUE is associated with a .018 × 1.40 = .025 bigger decline in the annual growth rate of total mortgages during the crash. This is over 32% of the average annual decline in total mortgage originations during this period.

So the results suggest that low local firm supply MSAs experienced a substantially larger increase in total mortgages during the bubble and a bigger decline during the crash. It is important to remember that because we include state effects in the regression specifications, we are not identifying these estimates by comparing MSAs across large geographic areas. For example, we are not using differences in the behavior of MSAs in Sand States compared to other MSAs located in other parts of the country to identify these estimates. With the state effects, we are comparing MSAs located in the same state and measuring whether the MSAs with lower RESIDUAL BOOK VALUEs have bigger changes in mortgage origination.

In columns (3) and (4) of Panel A of Table 5, we show the results of similar regressions presented in the first two columns except the dependent variable is now the annual percentage change in mortgage origination of second homes. The main story of the results is very similar to the results for total mortgages. Second home mortgages grow more in low local firm supply MSAs during the boom and fall more during the bust. The magnitudes of the coefficients are also similar to before. The coefficient on RESIDUAL BOOK VALUE in in column (3) suggests that a one standard deviation decrease in RESIDUAL BOOK VALUE increases the annual growth rate of second home mortgages during the bubble years by about 13%. A similar calculation for the crash period suggests that a one standard deviation decrease in RESIDUAL BOOK VALUE is associated with a 6% lower decline in second mortgages.

In Panel B of Table 5 we run the same regressions as presented in Panel A, but we restrict our sample to only the large MSAs. The basic story for the large MSAs is almost identical to the conclusions from the entire sample. In fact, the implied magnitudes of the low local firm supply on mortgages are typically larger using only the big MSAs. So our results are not being driven by tiny cities.

We next turn to examining the relationship between RESIDUAL BOOK VALUE and home price appreciation. The regression specification is almost identical to our specification for mortgage origination, but now the dependent variable is the average annual change in home prices across MSAs.

The results are presented in Table 6. Panel A shows the results using all MSAs. Panel B of Table 6 shows the results for only the large MSAs. The results are similar to Panel A. So we will focus our discussion on Panel A. Column (1) uses the bubble sample period of 2000-2006. The coefficient on RESIDUAL BOOK VALUE is negative and statistically different than zero (t-statistic = -2.2), indicating that MSAs with low local stock supply experienced a larger growth in housing prices during the boom. The magnitude of the coefficient indicates that a one standard deviation decrease in RESIDUAL BOOK VALUE increases the annual growth of housing prices by about $(.002 \times 1.40)/.09{=}3\%$ relative to the mean price growth over this time.

The coefficient on the housing supply elasticity is the expected sign; more elastic MSAs experienced a smaller increase in housing prices during the boom. MSAs with faster growing total income in the 1990s experienced a smaller increase in housing prices during the bubble period. The coefficient on the 1990s MSA high population growth dummy is positive but small.

We hypothesize that the effect of RESIDUAL BOOK VALUE on the growth of house prices in a MSA should depend on the supply elasticity of housing in that MSA. If the supply is very elastic, there should probably be little effect of local stock supply on prices. In low stock supply MSAs where the supply elasticity is very high, the market can adjust to the higher demand for housing with little change in price. So we want to break up the effect RESIDUAL BOOK VALUE to see whether the effect is concentrated among MSAs with lower supply elasticities.

These results are shown in column (2) of Table 6. We add to the regression specification dummies for the quintile of supply elasticity of the MSA and the interaction of these supply elasticity quintiles with RESIDUAL BOOK VALUE.[10] The coefficient on RESIDUAL BOOK VALUE is again negative and statistically different than zero. The magnitude of the coefficient is substantially bigger in absolute value than the result in column (1), indicating that the effect of RESIDUAL BOOK VALUE for MSA in the bottom quintile of the supply elasticity distribution is substantially bigger than the effect for the average MSA. The size of the coefficient suggests that a one standard deviation decrease in RESIDUAL BOOK VALUE increases the annual growth of housing prices by about $(.008 \times 1.40)/.09=12\%$. The coefficients on the interactions of RESIDUAL BOOK VALUE and elasticity quintiles are monotonically increasing with supply quintile. The effect of residual gets smaller as supply elasticity gets larger; the coefficients on the interactions suggest that the effect of RESIDUAL BOOK VALUE is basically zero for MSAs in the top two elasticity quintiles.

Columns (3) and (4) of Table 6 show the same specifications of the first two columns

[10]The omitted elasticity quintile is the lowest. That is, MSAs with the lowest supply elasticity.

except that the sample includes the bust years. Column (3) shows that low local firm supply MSAs have a larger decline in prices during the bust compared to other MSAs. The magnitude of the coefficient suggests that a one standard deviation decrease in RESIDUAL BOOK VALUE is associated with a (.003 × 1.40)/.06=7% bigger decline than the average.

The results in column (4) again show that this average effect of RESIDUAL BOOK VALUE is concentrated among MSAs with low supply elasticities. For the most inelastic MSAs, a one standard deviation decrease in RESIDUAL BOOK VALUE is associated with a (.008 × 1.40)/.06=19% bigger decline than the average. Just like the results for the boom period, RESIDUAL BOOK VALUE does not have much of an effect on MSAs in the highest two supply elasticity quintiles.

We can present these housing price appreciation results graphically in Figure 2. The figure shows the time-series of average home price appreciation by the groups of elasticity and RESIDUAL BOOK VALUE. We sort the MSAs into three housing supply elasticity groups (Lowest Elasticity 1/3, Middle Elasticity 1/3 and Highest Elasticity 1/3).[11] Then independently, we sort the MSAs into a low versus high RESIDUAL BOOK VALUE groups. The low RESIDUAL BOOK VALUE group is in Blue and the high RESIDUAL BOOK VALUE group in Red.[12]

First, notice that in the highest housing supply elasticity MSAs, there is not a substantial difference in the price appreciation of low and high local stock supply MSAs over the 2000-2010 period. This is consistent with Glaeser et al. (2008) who show that housing supply elasticity ought to dampen down any effects of speculative demand on prices.

We next consider the lowest housing supply elasticity MSAs. We see here that there are now significant deviations between the blue low RESIDUAL BOOK VALUE MSAs and the red high RESIDUAL BOOK VALUE MSAs. There is more of a cycle for the MSAs where real estate is the only game in town. The middle elasticity MSAs have a similar cycle as the

[11] We present the results with elasticity terciles to make the figure easier to read. As with the regressions, we can split the MSAs into housing elasticity quintiles and obtain very similar results.

[12] In Panel B of Figure 2, we use the RATIO measure of Hong et al. (2008) instead of RESIDUAL BOOK VALUE. As can be seen, we find very similar results using either measure.

low elasticity MSAs.

2.4. Explaining Sand States with RESIDUAL BOOK VALUE

We have shown that MSAs in Sand States are more likely to have low local stock supply. Also, we have shown that MSAs with low local stock supply are more likely to have had bigger increases in housing prices and mortgage origination during the housing boom and that pattern turns around during the bust. So the obvious question to ask is how much of the boom/bust pattern we see in the Sand States during the 2000s can be explained by the fact that those MSAs are more likely to have low local stock supply.

To isolate this effect, our strategy is to use our previous estimates of the effect of RESIDUAL BOOK VALUE on mortgage origination and housing price appreciation to predict how mortgage origination and housing price appreciation would have changed over the 2000s if there were no difference across MSAs in RESIDUAL BOOK VALUE. For example, we estimate the effect of RESIDUAL BOOK VALUE on price appreciation using our model shown in column (2) of Table 6. Using the coefficients of this model, we then predict what the housing price appreciation would have been if every MSA had the same value of RESIDUAL BOOK VALUE.[13] We can then regress this predicted housing price appreciation variable on a dummy for the MSA being in a Sand State to see if the coefficient on this dummy is substantially different than when we estimate a similar regression using the actual price appreciation of the MSA.

The results of this exercise are presented in Table 7. Column (1) shows the estimates of the actual percentage change of total mortgage origination on a Sand State dummy for the boom years. The coefficient on the Sand State dummy is positive and large. The coefficient suggests that being in a Sand State increases the growth of mortgage origination during the boom by about .08/.27=33%. This is an annual figure. So over the 7 years of the boom, this implies an over 200% higher increase in Sand State MSAs than other MSAs. This striking

[13] We set the value to zero for every MSA.

magnitude already controls for supply elasticity, the high income growth dummy, and high population growth dummy.

In column (2), we present the results of the same regression, except the dependent variable is the predicted growth in mortgage origination of MSAs if their RESIDUAL BOOK VALUE equals zero.[14] The coefficient on the Sand State dummy is now substantially smaller. The difference suggests that the only-game-in-town effect can explain about 45% of the greater mortgage origination growth of MSAs in Sand States during the boom.

Columns (3) and (4) show the results of the same exercise for the bust years. The coefficient on the Sand State dummy in column (3) is negative and large in absolute value. Sand state MSAs experienced about a -.09/.08=113% larger decline in mortgage originations than the average decline. The coefficient on the Sand State dummy in column (4) indicates that the Sand State effect would have been 49% smaller without the only game in town effect.

Columns (5) through (8) of Table 7 show the same set of results using measures of growth in second home mortgages instead of total mortgages. The differences between the baseline Sand State effect and the predicted effect if there was no only-game-in-town effect are substantial. The results suggest that differences in RESIDUAL BOOK VALUE can explain all of the differences in the growth of second mortgages in MSAs located in Sand States during the 2000s.

Finally, in columns (9) through columns (12), we show how the effect of a MSA being located in a Sand State on price appreciation changes once we control for the only game in town effect. The coefficient on the Sand State dummy in column (9) indicates that Sand States experienced on average a .06/.09=67% higher price appreciation during the boom than other MSAs. The results on column (4) show that after removing the price appreciation differences across MSAs driven by differences in RESIDUAL BOOK VALUE, the Sand State dummy is 15% smaller.

[14] The coefficients on the models used to make these predictions of mortgage growth are presented in Table A2. These models allow the effect of RESIDUAL BOOK VALUE to vary across supply elasticity quintiles.

We find similar results for price appreciation during the bust. Unconditionally, Sand State MSAs had on average a .08/08=100% larger decline in home prices during the bust than other MSAs. However, after removing the variation associated with RESIDUAL BOOK VALUE, this extra decline in Sand States is 13% smaller.

This exercise indicates that a substantial part of the excess growth of mortgages and prices in Sand States during the housing boom and the excess declines of those measures in the bust can be attributed to the fact that sand state MSAs on average have low local stock supply. Almost half of the excess movement of total mortgage origination (and all of second home mortgage origination volatility) in the Sand States during the 2000s can be explained with an only-game-in-town effect. About 14% of the price volatility can be similarly explained.

Panel B of Table 7 presents the results for large MSAs. Here it is interesting to note that our effects are nearly double that of the all MSAs sample. For instance, for total mortgage origination, accounting for RESIDUAL BOOK VALUE leads to a decline in the Sand State dummy of 45% in Panel A. The analogous number is a decline of 119% in Panel B. That is, among large MSAs, the only-game-in-town effect virtually wipes out the Sand States effect. The same is true for investment home mortgage origination. For the price appreciation from 2000-2006, the magnitude goes from -15% to -84%.

2.5. Results Using RATIO Instead of RESIDUAL BOOK VALUE

For large MSAs, we can, as explained in the Introduction, use RATIO following Hong et al. (2008).[15] In the RATIO, there is much more variation in the book value per capita across MSAs than there is in the income per capita. This is consistent with the point in Hong et al. (2008) that RATIO is essentially variation in book per capita.[16]

[15] Hong et al. (2008) mostly use RATIO calculated at the Census Division and state level. We will concentrate on the effect of RATIO only at the MSA level.

[16] In particular, they find that RATIO is highly persistent and unassociated with the economic growth prospects in area. Their focus is on RATIO at the state level; in our empirical designs, given even closer proximity of real estate investments, we focus our analysis at the MSA level.

Figure 3 shows the geographic distribution of RATIO across the US. There is wide variation across MSAs in their RATIO value; within every region of the country, there are high and low RATIO MSAs. However, there are some geographic patterns. RATIO tends to be higher in the Northeast, for example, where lots of firms are headquartered. The West tends to have many of the lowest RATIO MSAs. In our empirical work, we will make sure that our results are not being solely driven by comparisons of MSAs across regions of the US.

We run our models presented in Tables 5 and 6, showing how RATIO explains mortgage originations and price appreciation using RATIO instead of RESIDUAL BOOK VALUE. Table 8 shows these results for the sample of large MSAs. Columns (1) and (2) examine total mortgage origination. The coefficient on log RATIO in column (1) suggests that low local stock supply MSAs (low RATIO MSAs) experienced a larger increase in mortgages during the housing boom. The size of the coefficient suggests that a one standard deviation decrease in log RATIO increased the price appreciation of a MSA during the boom by about $(.038 \times 1.00)/.28=13\%$ of the mean. In column (2), there is a similarly strong relationship between log RATIO and total mortgage origination during the crash. A one standard deviation decrease in log RATIO increases the decline in mortgage origination by about $(.045 \times 1.00)/.08=56\%$ relative to the mean decline. The results presented for second mortgage origination in columns (3) and (4) show the same pattern.

The last two columns present the results for price appreciation during the boom and crash. From column (5), a one standard deviation decline in log RATIO increases the price appreciation of a MSA during the bubble by about $(.009 \times 1.00)/.10=9\%$ relative to the mean increase. Similarly, a one standard deviation decrease in log RATIO increases the price decline during the bust by about $(.011 \times 1.00)/.06=18\%$ relative to the mean decline. So the results indicate that using RATIO as a measure of local stock supply in large MSAs tells a very similar story to using RESIDUAL BOOK VALUE.

3. Household Portfolio Tilt Toward Investment Homes

Having established the importance of the local supply of equities for the Sand State effect, we turn to identifying the mechanism at the heart of these results. Namely, in low local supply of equities MSAs, households turn to buying investment homes as opposed to stocks, thereby creating the amplification effect that we show previously can explain the Sand States effect. To do this, we bring to bear new data from the Survey of Consumer Finance.

3.1. Data from the SCF

As a survey of household finances and wealth, the Survey of Consumer Finance (SCF) includes some assets that are broadly shared across the population (e.g. bank savings accounts) as well as some that are held more narrowly and that are concentrated in the tails of the distribution (e.g. direct ownership of bonds). To support estimates of a variety of financial characteristics as well as the overall distribution of wealth, the survey employs a dual-frame sample design.

A national area-probability (AP) sample provides good coverage of widely spread characteristics. The AP sample selects household units with equal probability from primary sampling units that are selected through a multistage selection procedure, which includes stratification by a variety of characteristics, and selection proportional to their population.

Because of the concentration of assets and non-random survey response by wealth, the SCF also employs a list sample which is developed from statistical records derived from tax returns under an agreement with the IRS.[17] (See Kennickell (2000) for additional details on the SCF list sample.) This list sample consists of households with a high probability of having high net worth.[18]

[17]See Wilson and William J. Smith (1983) and Internal Revenue Service for a description of the SOI file. The file used for each survey largely contains data from tax returns filed for the tax year two years before the year the survey takes place. See Kennickell (1998) for a detailed description of the selection of the 1998 list sample.

[18]For reasons related to cost control on the survey, the geographic distribution of the list sample is constrained to that of the area-probability sample.

The SCF joins the observations from the AP and list sample through weighting.[19] The weighting design adjusts each sample separately using all the useful information that can be brought to bear in creating post-strata. The final weights are adjusted so that the combined sample is nationally representative of the population and assets. These weights are used in all regressions.

3.2. Merging RATIO with SCF

For our analysis, we focus on the 65 MSAs with population larger than 750,000. Our independent variable of interest is log RATIO. We work with logs because it offers a better fit in the context of household portfolios in contrast to our cross-MSA regressions from before. Due to a few MSAs with RATIO close to zero, we add 0.00001 to RATIO before we take logs. This scaling is arbitrary and does not materially affect our conclusions. RATIO is available at the MSA-level and is merged into the SCF data in the following way. A new MSA variable is created in the SCF using a state and county FIPS to MSA correspondence consistent with the geography used in calculating RATIO. The RATIO variable, for the years 1995, 1998, 2001, 2004, 2007, and 2010, is merged into the SCF data by MSA and year.[20]

Ultimately, there are around 18,000 households sampled by the SCF between 1995 and 2010 that are used in this portion of the analysis. Because of the multiple imputation process (five implicates to generate a distribution for the imputed values) for missing values, there are 89,956 household-level records in the data. Standard errors in the SCF regressions are based on weighted data, and also are adjusted for the multiple implicates.

[19]The evolution of the SCF weighting design is summarized in Kennickell (2000), with additional background by Kennickell and Woodburn (1992).

[20]Starting in the 2001, the public-use version of the SCF does not include any geographic identifiers. Prior to 2001 the public-use version only included the very broad 9-level Census Division code. There is an internal version of the survey, however, that is available only to the Board of Governors economists working on the SCF that has MSA-level (and lower levels of geography including zip-code and county) identification of where the household resides. These geographic variables were used to merge the RATIO variable into the SCF.

3.3. Definitions of Investment Real Estate and Stock Holdings

A group of dependent variables are represented as shares: they sum the dollar value of the household's portion of all investment real estate, vacation home, or directly-held stocks, and divide by total assets. The SCF clearly identifies primary residences and second homes. Investment real estate is aggregated from the following sub-categories given by the SCF in the "property type" variable (x1703, x1803, and x1903): code 11 (land only: lot, tract, acreage; building lots,"farmland"); code 13 (substantial land and other type of structure); code 15 (recreational property; sports field; golf course); code 24 (mobile home park); code 40 (one single family home); code 41 (multiple single family homes); code 42 (duplex 2 unit residence); code 43 (triplex); code 44 (fourplex); code 45 (5 or more); code 46 (apartment house, units unknown, rental units, or property nfs); code 47 (other business commercial property); code 48 (business/commercial and residential combination); code 49 (condo; co-op); and code 50 (residential, nec.).[21]

% INVESTMENT REAL ESTATE is then the dollar value of investment real estate divided by the household's total assets. Vacation homes are aggregated from the following categories reported by SCF: code 21 (seasonal/vacation); code 25 (time share); code 12 (substantial land and seasonal or other residence); and code 999 (other vacation home mapped from the mop-up). Analogously, we calculate % VACATION HOME. We also have data on how much households own of stocks. We calculate % DIRECTLY-HELD STOCKS as the fraction of directly-held stock holdings in a household's total asset.

We report in Table 9 the summary statistics of the SCF for the 1995, 1998, 2001, 2004,

[21] As a robustness check of our results, we consider different definitions of investment real estate. First, we define investment real estate as strictly residential. Specifically, we consider the following subset of codes: code 24 (mobile home park); code 40 (one single family home); code 41 (multiple single family homes); code 42 (duplex 2 unit residence); code 43 (triplex); code 44 (fourplex); code 45 (5 or more); code 46 (apartment house, units unknown, rental units, or property nfs); code 48 (business/commercial and residential combination); code 49 (condo; co-op); and code 50 (residential, nec.). Second, the SCF asks whether the household or Primary Economic Unit (PEU) earns any money (and how much) from other real estate. This definition of investment real estate simply takes the asset value for those other properties that are generating income for the PEU. We find that both definitions show similar results to our original definition above.

2007, and 2010 waves. In our sample of households, they are more apt to have an investment real estate than a vacation home. % INVESTMENT REAL ESTATE, the fraction of investment real estate in household's total asset, has a mean of 0.03 with a standard deviation of 0.10. % VACATION HOME, the fraction of vacation home in household's total asset, has a mean of 0.01 with a standard deviation of 0.04. % DIRECTLY-HELD STOCKS (as opposed to held through intermediated vehicles such as mutual funds), the fraction of directly-held stock holdings in household's total asset, has a mean of 0.02 with a standard deviation of 0.08.

In addition to these share variables, we also create related dummy variables to capture whether households have any investment real estate, vacation home or directly-held stocks. HAVE INVESTMENT REAL ESTATE equals one if the household has any investment real estate and zero otherwise. The mean is 0.13 with a standard deviation of 0.34. So 13% of our households own some investment real estate. HAVE VACATION HOME equals one if the household has a vacation home and zero otherwise. The mean is 0.06 with a standard deviation of 0.24. HAVE DIRECTLY-HELD STOCKS equals one if the household has any directly-held stock and zero otherwise. The mean is 0.21 with a standard deviation of 0.41. Overall, our households are more likely to have stocks than investment homes and more likely to have investment homes than vacation homes.

In Table 9, we also report summary statistics for our right-hand side variables. Log RATIO is the log ratio of the total book value of firms headquartered in a MSA to the income in that MSA. The mean of log RATIO is -1.13 with a standard deviation of 1.08. MSA Unemployment Rate is unemployment rate from the Bureau of Labor Statistics. The mean is 5.65 and standard deviation is 2.26. MSA Home Price Index is the Federal Housing Finance Agency (FHFA) Housing Price Index at the MSA level and has a mean of 163 with a standard deviation of about 54. MSA Housing Affordability Index is the Housing Affordability Index from the National Association of Realtors.[22] The Housing Affordability

[22]http://www.realtor.org/topics/housing-affordability-index

Index measures whether or not a family with median income could qualify for a mortgage on a median-priced home. A higher index value indicates that a median-priced home is more affordable to a median income family. Since the index is available only for 2009 to 2011, we backfilled the data using time-series average of available Housing Affordability Index. The Housing Affordability Index has a mean of 120 with a standard deviation of about 58.

log HOUSE PRICE ZIP CODE is a home price index at zip code level from Zillow and has a mean of 12.22 with a standard deviation of about 0.64. We mainly use FHFA MSA home price index in our analysis, but we use Zillow home price index when we control for home prices at the zip code level since FHFA data are only available at the MSA level.[23]

FAMILY SIZE is the number of people in each Primary Economic Unit. Our households have on average around 2.44 members with a standard deviation of 1.42. Log HOUSEHOLD INCOME has a mean of 10.87, which is around 53 thousand dollars. The SCF Survey is ideal for our study because it includes higher net worth households who are most likely to be able to purchase an investment real estate. HOUSEHOLD NET WORTH has a mean of about 600,000 dollars with a standard deviation of 4,100,000. UNATTACHED FEMALE is an indicator that the family is headed by an unattached woman. About 27% of our samples are headed by an unattached woman. Table 2 also reports the breakdown of AGE of the head of each household. The sample is meant to be nationally representative and as such we see a distribution across the age cohorts. We also break down RACE. 68% of the households are white and 16% are black. Hispanics account for 10% of the households and Asians or others account for the remaining 5%.

We also break down EDUCATION of the head of each household. 13% have less than a high school education. 28% have high school or GED equivalent. 25% went to some college but do not have a Bachelor's degree. 21% have a Bachelor's degree. 7% have a Master's degree and 6% have an advanced degree such as PhD, JD, MD, or MBA. FAMILY STRUCTURE breaks down situations like whether or not the couple is married and has

[23] In the last part of our analysis, we also use Zillow prices in conjunction with monthly dollar rent by county from US Department of Housing and Urban Development to construct house price-to-rent ratios.

children. Here LWP stands for the “living with partner”.

3.4. Average FICO Scores by Zip Code

Because we know in the SCF the zip code of residence of each household, we can match the average FICO score of the residents of the household's zip code to the other information we have about the household. Similar to previous studies such as Mian and Sufi (2009), we divide zip codes into quartiles based on their average FICO scores. Table 10 presents the range of average FICO scores by quartile. Subprime borrowers are usually defined as having FICO scores below 660. The lowest quartile of zip codes have average FICO scores well below the subprime cut-off. For the second lowest quartile, the mean of the zip code FICO measures is only slightly above the subprime cut-off. Only in the top half of zip codes in the average FICO distribution is the typical household well above subprime. For much of our analysis, we will compare the results of households living in the top and bottom quartile of the zip code FICO distribution. Table 10 shows that we are then comparing households living in neighborhoods with predominately prime borrowers with households living in neighborhoods with predominantly subprime borrowers. Households in the SCF sample are roughly evenly distributed across the four FICO quartiles: 16% in the lowest FICO score group, 29% in the second group, 32% in the third group and 23% in the highest FICO score group.

3.5. Demographics and Portfolio Tilt between Investment Homes and Stocks

We begin in Table 11 by reporting the household-level panel logit regression results, measuring how household demographic characteristics affect dummy variables for investment real estate and stock ownership. In our baseline specification, we estimate a logit model of the asset holding indicator on the following household characteristics: log RATIO, HIGH FICO ZIP CODE INDICATOR, log HOUSEHOLD INCOME, log HOUSEHOLD NET WORTH, UNATTACHED FEMALE, AGE, RACE and EDUCATION. Also, included in all the logit specifications are year effects and a series of other household level controls for FAMILY STRUCTURE, and a linear and square term of FAMILY SIZE. For brevity, we do not report these coefficients on FAMILY STRUCTURE, and a linear and square term of FAMILY SIZE.

In column (1), the dependent variable is the dummy variable HAVE INVESTMENT REAL ESTATE. The coefficient in front of log RATIO in column (1) is -0.0751 with a t-statistic of -1.91. Households are less likely to own an investment real estate if they live in a high RATIO MSA with lots of stocks headquartered there. The marginal effect is -0.0074. This suggests that a one standard deviation increase in log RATIO (1.08) decreases the probability of a household owning investment real estate by about 0.0080 (.0074 $\times$ 1.08). Given that the unconditional probability is around 0.13, this is about a 6% decline in the probability.

Also, households living in high average FICO zip codes are less likely to hold investment real estate than other households (i.e. households in a low average FICO zip code).[24] The coefficient on High FICO Zip Code indicator is -0.309 with a t-statistic of -4.11. All else equal, households in prime zip codes are much less likely to own investment homes than those in higher-risk or subprime zip codes.

Not surprisingly, higher income, wealth and age substantially increase the likelihood of owning investment real estate. Being an unattached female substantially lowers the probability of owning a home. Non-white households are more likely to hold investment real estate (though only the dummy for being Black is statistically significant), conditional on income and wealth.

There are many potential reasons for why households in high FICO zip codes and white households hold less investment homes (and as we discuss in columns (3)-(4) below simultaneously hold more stocks) compared to those in low FICO zip codes and non-white households,

[24]Households in the top 25% of the FICO score distribution are classified as living in low risk zip codes. Households in the bottom 25% of the FICO score distribution are classified as living in a high risk zip code.

respectively. This heterogeneity in preferences for homes versus stocks can be traced from the household finance literature to investor sophistication, familiarity with homes either due to employment such as construction, or other social interaction and cultural biases. We take this underlying heterogeneity in preferences as given and will use it to our advantage to measure our only-game-in-town effect.

Toward this end, we need to consider potential confounders that muddy the interpretation of the relationship between log RATIO and household investment real estate decisions. As seen in Figure 3, there is some spatial correlation in RATIO across the US. RATIO tends to be higher in MSAs in the Northeast and lower in the West and the South. MSAs in the Midwest are likely to be in the middle of the RATIO distribution. We are worried that there might be different types of investors living in Northeast MSAs than other parts of the country. In other words, we see for instance that certain types of households (non-white, living in low FICO zip codes) prefer investment homes to stocks. To the extent that different MSAs might have different proportions of these households, we might incorrectly interpret log RATIO as an only-game-in-town effect.

Much of our empirical work will involve trying to rule out this alternative explanation for the relationship between log RATIO and investment behavior. To attempt to address such unobservables, we first try to identify the relationship between log RATIO and investment real estate behavior without using cross-region variation in log RATIO. In column (2), we re-estimate these relationships but now also include Census Division $\times$ Year effects in the logit specification. Therefore, comparisons of the behavior of households in Northeast MSAs to households in the South are not being used to estimate the propensity of households to hold investment real estate. All our results from column (2) are similar to those from column (1). The coefficient on log RATIO in column (2) is now -0.0695 and with a t-statistic of -2.37. A one standard deviation increase in log RATIO decreases the probability that a household holds investment real estate by again about 6% of the unconditional mean, almost the same as in column (1). Therefore, our results are not being driven by across Census region comparisons of MSAs.

Both of these logit specifications indicate that households in MSAs with a higher log RATIO are less likely to hold investment real estate. Panel A of Figure 4 shows a residual scatter plot of the relationship between investment real estate holding and log RATIO at the MSA/year level. The regression specification is the same as the one shown in column (1) of Table 11. Residualized log RATIO averaged at the MSA/year level is plotted against residualized investment real estate holding propensity again at the MSA/year level. The scatter plot shows how the negative relationship between the two variables is measured.

For the rest of the results shown in Table 11, we examine the relationship between log RATIO and the propensity for households to directly hold stocks. Recall that our hypothesis suggests that in MSAs with fewer stocks nearby, households are also less likely to have directly-held stocks in their portfolio. The specifications in columns (3)-(4) are identical to the first two columns of Table 11 except that the outcome variable is instead an indicator for directly-held stocks.

Consistent with an only-game-in-town effect, we find in both of our specifications a positive and statistically significant relationship between log RATIO and the propensity to hold stocks directly. The coefficient of interest in column (3) is 0.126 with a t-statistic of 4.01. A one standard deviation increase in log RATIO is associated with an 8% increase in the probability that a household holds stock relative to the unconditional probability. Households from high FICO zip codes are also more likely to own stock. So are households with higher income and net worth and younger households. Households with an unattached female as head are less likely to own stock. Non-whites are less likely to own stock, whereas more educated households are more likely to own stock. These latter two findings are consistent with Hong et al. (2004). Removing across-region variation in log RATIO by adding Census Division by year effects in the specification in column (4) again does not change the coefficient on log RATIO nor the coefficients on the other covariates.

One might worry that our findings in columns (3)-(4) of Table 11 are somehow hardwired. If we know that an MSA has fewer stocks, does not that mean that households there necessarily are less likely to be tilted toward stocks? This is not the case. For instance, it is possible that an MSA with a lower supply of equities could actually have all of its shares held by locals, giving us a negative rather than a positive coefficient of interest.

Having examined just the extensive margin, i.e., the decision to own or not own investment homes and stocks, we now turn to examining how log RATIO affects the total amount of investment real estate and stocks a household holds measured as a percentage of the household's portfolio devoted to those two assets. Table 12 presents these results. It is identical to Table 11 in structure except for two changes. In columns (1)-(2), the dependent variable is the percentage of a household's portfolio devoted to investment real estate instead of an indicator for holding investment real estate. Similarly, in columns (3)-(4), the dependent variable is the percentage of a household's portfolio devoted to stocks. Because the dependent variables in Table 12 are fractions bounded between 0 and 1, we estimate the models using tobits instead of logits.

In columns (1)-(2) of Table 12, the relationship between log RATIO and the amount of households' portfolios devoted to investment real estate is again similar to the logit results. The coefficient of interest in column (1) is -.0144 with a t-statistic of -2.11. As with the logit results, the magnitude of the coefficient on log RATIO is unaffected by the inclusion of Census Division by year effects in column (2). The coefficients on log RATIO in columns (1)-(2) imply that the effect of a one standard deviation increase in log RATIO on the percentage of assets in investment real estate ranges from a decrease of 40 to 52% relative to the mean. Panel B of Figure 4 shows the residualized scatter plot of the relationship between log RATIO at the MSA/year level and the residualized average percent holding of investment real estate again at the MSA/year level.

The coefficients on the other demographic characteristics are again very similar to the logit results reported in Table 11. Households living in higher average FICO zip codes have less of their portfolio devoted to investment real estate. The coefficient of interest in column

(1) is -0.0551 with a t-statistic of -3.96. A similar coefficient obtains in column (2) when we introduce Census Division × year effects.

In columns (3)-(4), we show there is a positive and statistically significant effect of log RATIO on the proportion of a household's assets devoted to stocks. The coefficient of interest in column (3) is 0.015 with a t-statistic of 4.5. The tobit coefficients suggest that a one standard deviation increase in log RATIO raises a household's holding of stocks about 80% relative to the mean. Households living in high FICO zip codes own more stocks and the economic effect is also large.

In sum, we have established three novel patterns in Tables 11 and 12. First, that households in high log RATIO MSAs are, all else equal, less likely to own investment homes and more likely to own stocks. Second, households in high FICO zip codes tilt away from real estate investment homes and towards stocks compared to their low FICO zip code counterparts. Third, white households tilt towards stocks as opposed to investment real estate compared to their non-white counterparts.

3.6. Interacting log RATIO and High FICO Zip Code Indicator and Including MSA × Year Effects

Even with Census × Year Effects, we might still worry that there is significant variation across MSAs within the same Census Division in the type of households living in these MSAs. In other words, we know that certain types of households prefer investment homes to stocks. To the extent that different MSAs, even if they are in the same Census Division, might have different proportions of these households in a way that is correlated with log RATIO, we might incorrectly interpret log RATIO as an only-game-in-town effect.

To address this concern, we estimate an empirical model that measures the behavior of households within a MSA. To set up this specification, first recall that households in the top 25% of the FICO score distribution are classified as living in low risk zip codes. Households in the bottom 25% of the FICO score distribution are classified as living in a high risk zip code. We leave out the middle 50% of the FICO distribution to make sure we are comparing households across a wide range of average FICO scores. Then, within each MSA, we can calculate the difference in propensities to buy investment homes between households living in high FICO zip codes and low FICO zip codes and then see how this difference varies across MSAs by log RATIO. There will be somewhat different proportions of these households across MSAs, but the key is within every MSA we have some investors in both groups.

We expect that in low RATIO MSAs even households in high FICO zip codes will tilt toward investment homes because there are few local stocks for them to invest in. That is, in MSAs with few local firms, the behavior of households in low FICO and high FICO zip codes ought to be more similar than in MSAs with many local firms. Everyone, even high FICO households who might on average have a preference for stocks as opposed to investment homes, is tilted toward investment homes in MSAs with few local firm if our only-game-in-town effect holds.

To see if this is the case, we estimate models where an indicator for having investment real estate is regressed on log RATIO, an indicator for living in a high FICO score zip code, an interaction of log RATIO and the high FICO score zip code indicator and other controls. The coefficient of interest is on the interaction term; it tells us whether the difference in behavior between households in high versus low FICO zip codes varies with log RATIO in the way we predict.

One of the benefits of this difference-in-differences specification is that we can include MSA by year effects in the model and still identify the interaction term. This is because there is variation in the interaction term within MSA by year cells because some households live within low and high risk zip codes within the same MSA at the same time. Therefore, we can difference out MSA-level differences in investor preferences.

Columns (1) and (2) of Table 13 show the results with an indicator for having investment real estate as the dependent variable.[25] In column (1), we present the model without the

[25]We use linear probability models for columns (1) through (2) of Table 13 instead of logits. The interaction terms are easier to interpret in linear models. If we estimate the logits presented earlier as linear probability

interaction term to show that we obtain the same negative relationship between log RATIO and the probability of holding investment real estate even for this smaller sample of households living in high and low FICO score zip codes. Notice that the coefficient on living in a high FICO zip code is negative and similar in magnitude to our earlier results. The coefficient, -.0087 with a t-statistic of -1.38, is comparable in economic magnitude to the implied economic effect from the logit specification in column (1) 1 of Table 11. However, the coefficient is more imprecise because we have omitted roughly half of our sample.

In column (2), we add the interaction term and MSA by year fixed effects to the regression specification. The coefficient of interest is -.0218 with a t-statistic of -2.14. The value of the interaction term indicates that the negative sensitivity of investment real estate holding to log RATIO is larger for households in high FICO zip codes. In fact, given that the magnitude of the interaction term (-.0218) is almost three times larger than the average effect of log RATIO for the entire sample (-.0087) and that the fraction of low and high FICO households are roughly comparable, we can conclude that almost all the average effect of log RATIO on owning investment real estate is coming from the sensitivity of high FICO households' investment behavior to log RATIO.

Columns (3) and (4) show similar results for the tobit models. The basic relationship between log RATIO and real estate investment holdings in column (3) is almost the same as previously shown in Table 12. Notice that the coefficient of interest here is -.0111 with a t-statistic of -1.33, which is almost identical to the corresponding coefficient of -.0144 from column (1) of Table 12. Again, the coefficient here is less precise than before because we have dropped roughly half the sample in drawing a comparison between the extreme low and high FICO groups. The interaction term in column (4) is -.0408 with a t-statistic -14.57. It is about four times larger than the average effect of log RATIO on the percentage of investment home as a fraction of total assets. So we can conclude that similar to the extensive margin, the effect of log RATIO on the total portfolio weight in investment homes is driven by high

models instead, we find very similar quantitative and qualitative results.

FICO households.

We have shown before that households living in high FICO zip codes want to hold less investment real estate and more stocks than other households. But in MSAs with few local stock investment opportunities (low log RATIO), households living in high FICO zip codes look more like households living in low FICO zip codes. They do not move away from investment real estate as much as households in MSAs with many local stock investments. In fact, our regression estimates allow us to calculate how much of a movement in log RATIO is necessary to offset this level difference in the propensity of households in high and low FICO neighborhoods to hold investment real estate. Our estimates suggests that a one standard deviation (1.08) decrease in a MSA's log RATIO increases the probability that households in high FICO zip codes hold investment real estate enough to offset the level difference between households in low and high FICO zip codes.

3.7. Robustness Checks

We have completed several analyses to show the robustness of our SCF results. For brevity, we will only briefly describe these exercises. In an Internet appendix, we will describe these results more fully. First, we show that our FICO difference-in-difference results are robust to including all observations, not just households from the top and bottom FICO quartile. Second, we show that in a horserace, the interaction of log RATIO and High FICO is robust to the inclusion of other interactions of log RATIO with demographic controls such as education and race. Then we show that there is no relationship between log RATIO and whether households own vacation homes. We show the difference-in-difference results explain stock holdings and not just investment real estate.

Finally, we show that our difference-in-difference results are robust to controls for conditions in MSAs. First, we show that controlling for housing affordability in a MSA does not substantially change our results. Lastly, we show that controlling for MSA economic conditions in a variety of ways again does not change our conclusions.

4. Conclusion

We show that the excesses of the historic US housing cycle of the 2000s, which were concentrated in the Sand States and difficult to rationalize using existing factors, can be explained by an only-game-in-town effect. Sand State MSAs having an abnormally low supply of public equities (measured using the book value of firms headquartered there) relative to total income. This only-game-in-town factor can explain over 50% of the excess total mortgage originations, nearly all of the excess of investment home mortgage originations, and almost 20% of the excess price volatility of the Sand States.

We argue that the reason is that local biased households purchase investment homes nearby rather than owning stocks, thereby amplifying the housing cycle in these MSAs. We confirm this mechanism using data from the Survey of Consumer Finances. We establish that households residing in an MSA with few publicly traded firms headquartered there are more likely to purchase an investment home nearby. Households in these areas are also less likely to own stocks. We use a difference-in-differences estimation strategy to show that we are capturing our only-game-in-town effect as opposed to heterogeneity in preferences across MSAs.

References

Barber, B. M., and T. Odean. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21:785–818.

Chinco, A. M., and C. J. Mayer. 2012. Distant Speculators and Asset Bubbles in the Housing Market. *NBER Working Paper* pp. 1–52.

French, K. R., and J. M. Poterba. 1991. Investor Diversification and International Equity Markets. *American Economic Review* 81:222–226.

Glaeser, E. L., J. D. Gottlieb, and J. Gyourko. 2013. Can Cheap Credit Explain the Housing Boom? *Housing and the Financial Crisis, University of Chicago Press* pp. 301–359.

Glaeser, E. L., J. Gyourko, and A. Saiz. 2008. Housing Supply and Housing Bubbles. *Journal of Urban Economics* 64:198–217.

Grinblatt, M., and M. Keloharju. 2001. How Distance, Language, and Culture Influence Stockholdings and Trades. *Journal of Finance* 56:1053–1073.

Guiso, L., P. Sapienza, and L. Zingales. 2009. Cultural Biases in Economic Exchange? *Quarterly Journal of Economics* 124.

Haughwout, A., D. Lee, J. Tracy, and W. van der Klaauw. 2011. Real Estate Investor, the Leverage Cycle, and the Housing Market Crisis. *Federal Reserve Bank of New York Staff Reports 514* .

Hong, H., J. D. Kubik, and J. C. Stein. 2004. Social Interaction and Stock-Market Participation. *Journal of Finance* 59:137–163.

Hong, H., J. D. Kubik, and J. C. Stein. 2008. The Only Game in Town: Stock-Price Consequences of Local Bias. *Journal of Financial Economics* 90:20–37.

Huberman, G. 2001. Familiarity Breeds Investment. *Review of Financial Studies* 14:659–680.

Internal Revenue Service. 1992. *Individual Income Tax Returns, 1990.* Internal Revenue Service.

Kennickell, A. B. 1998. List Sample Design for the 1998 Survey of Consumer Finances. *FRB Working Paper* .

Kennickell, A. B. 2000. Wealth Measurement in the Survey of Consumer Finances: Methodology and Directions for Future Research. *FRB Working Paper* .

Kennickell, A. B., and R. L. Woodburn. 1992. Estimation of Household Net Worth Using Model-Based and Design-Based Weights: Evidence from the 1989 Survey of Consumer Finances. *FRB Working Paper* .

Keys, B. J., T. Mukherjee, A. Seru, and V. Vig. 2009. Financial Regulation and Securitization: Evidence from Subprime Mortgage Loans. *Journal of Monetary Economics* 56:700–720.

Li, W., and Z. Gao. 2013. Real Estate Investors and the Boom and Bust of the US Housing Market. *SSRN Working Paper* pp. 1–38.

Mian, A., and A. Sufi. 2009. The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis. *The Quarterly Journal of Economics* 124:1449–1496.

Olesiuk, S. M., and K. R. Kalser. 2009. The sand states: anatomy of a perfect housing-market storm. *FDIC Quarterly* 3:30–32.

Palmer, C. 2014. Why Did So Many Subprime Borrowers Default During the Crisis: Loose Credit or Plummeting Prices? *manuscript, MIT* .

Shiller, R. J. 2005. Irrational Exuberance, 2nd Edition. *Princeton University Press* .

Wilson, O. H., and J. William J. Smith. 1983. Access to Tax Records for Statistical Purposes. *Proceedings of the Section on Survey Research Methods, American Statistical Association* pp. 595–601.

Table 1: Characteristics of Investment and Vacation Home Purchases

Panel A reports the distance of second homes from primary residences from the *Investment and Vacation Home Buyers Survey* by the National Association of Realtors. The number of respondents for the Survey was 8205 for year 2004, 1034 for year 2005, 1965 for year 2007, 1924 for year 2008, 1930 for year 2009, 1895 for year 2010, 2241 for year 2011, 2326 for year 2012, and 2008 for year 2013. Panel B reports the number of home sales by intended use. The share in total home sales are reported in parenthesis.

Panel A : *Distance from Primary Residence*

<table>
<tr><th>Vacation Properties (%)</th><th>2013</th><th>2012</th><th>2011</th><th>2010</th><th>2009</th><th>2008</th><th>2007</th><th>2005</th><th>2004</th></tr>
<tr><td>5 miles or less</td><td>2</td><td>7</td><td>2</td><td>2</td><td>2</td><td>2</td><td>4</td><td rowspan="2">3</td><td rowspan="4">44</td></tr>
<tr><td>6 to 10 miles</td><td>4</td><td>4</td><td>3</td><td>3</td><td>4</td><td>2</td><td rowspan="2">2</td></tr>
<tr><td>11 to 15 miles</td><td>6</td><td>2</td><td>2</td><td>3</td><td>1</td><td>1</td><td rowspan="2">4</td></tr>
<tr><td>16 to 20 miles</td><td>6</td><td>5</td><td>4</td><td>4</td><td>3</td><td>4</td><td>4</td></tr>
<tr><td>21 to 50 miles</td><td>13</td><td>6</td><td>10</td><td>8</td><td>8</td><td>7</td><td>7</td><td>8</td><td rowspan="2">19</td></tr>
<tr><td>51 to 100 miles</td><td>15</td><td>10</td><td>14</td><td>11</td><td>16</td><td>19</td><td>18</td><td>19</td></tr>
<tr><td>101 to 500 miles</td><td>20</td><td>20</td><td>27</td><td>27</td><td>27</td><td>30</td><td>30</td><td>31</td><td>23</td></tr>
<tr><td>501 to 1000 miles</td><td>12</td><td>15</td><td>11</td><td>11</td><td>12</td><td>14</td><td>12</td><td>9</td><td>3</td></tr>
<tr><td>1001 miles or more</td><td>22</td><td>31</td><td>26</td><td>30</td><td>28</td><td>22</td><td>22</td><td>25</td><td>10</td></tr>
<tr><td>Median (miles)</td><td>180</td><td>435</td><td>305</td><td>375</td><td>348</td><td>316</td><td>287</td><td>220</td><td>49</td></tr>
<tr><th>Investment Properties (%)</th><th>2013</th><th>2012</th><th>2011</th><th>2010</th><th>2009</th><th>2008</th><th>2007</th><th>2005</th><th>2004</th></tr>
<tr><td>5 miles or less</td><td>15</td><td>24</td><td>17</td><td>20</td><td>18</td><td>24</td><td>16</td><td rowspan="2">52</td><td rowspan="4">69</td></tr>
<tr><td>6 to 10 miles</td><td>15</td><td>13</td><td>9</td><td>13</td><td>12</td><td>11</td><td>11</td></tr>
<tr><td>11 to 15 miles</td><td>10</td><td>6</td><td>12</td><td>8</td><td>9</td><td>8</td><td>8</td><td rowspan="2">17</td></tr>
<tr><td>16 to 20 miles</td><td>10</td><td>7</td><td>10</td><td>15</td><td>9</td><td>11</td><td>13</td></tr>
<tr><td>21 to 50 miles</td><td>12</td><td>15</td><td>13</td><td>12</td><td>11</td><td>8</td><td>15</td><td>6</td><td rowspan="2">9</td></tr>
<tr><td>51 to 100 miles</td><td>9</td><td>6</td><td>10</td><td>4</td><td>9</td><td>7</td><td>10</td><td>5</td></tr>
<tr><td>101 to 500 miles</td><td>10</td><td>10</td><td>10</td><td>10</td><td>11</td><td>12</td><td>12</td><td>8</td><td>5</td></tr>
<tr><td>501 to 1000 miles</td><td>8</td><td>8</td><td>7</td><td>5</td><td>7</td><td>8</td><td>6</td><td>3</td><td>3</td></tr>
<tr><td>1001 miles or more</td><td>12</td><td>11</td><td>13</td><td>13</td><td>14</td><td>11</td><td>10</td><td>9</td><td>14</td></tr>
<tr><td>Median (miles)</td><td>20</td><td>21</td><td>25</td><td>19</td><td>24</td><td>19</td><td>27</td><td>10</td><td>18</td></tr>
</table>

Panel B : *Home Sales by Intended Use*

	Primary Residences	Vacation Properties	Investment Properties
2003	4,844,000 (67%)	850,000 (12%)	1,572,000 (22%)
2004	5,106,000 (64%)	872,000 (11%)	2,003,000 (25%)
2005	5,023,000 (60%)	1,019,000 (12%)	2,317,000 (28%)
2006	4,817,000 (64%)	1,067,000 (14%)	1,646,000 (22%)
2007	3,925,000 (67%)	670,000 (12%)	1,221,000 (21%)
2008	3,207,000 (70%)	436,000 (9%)	951,000 (21%)
2009	3,441,000 (73%)	471,000 (10%)	801,000 (17%)
2010	3,294,000 (73%)	469,000 (10%)	749,000 (17%)
2011	2,785,000 (61%)	502,000 (11%)	1,233,000 (27%)
2012	3,268,000 (65%)	553,000 (11%)	1,207,000 (24%)
2013	3,697,000 (67%)	717,000 (13%)	1,104,000 (20%)

Table 2: Summary Statistics

Panel A reports the summary statistics for all MSAs. Log Total Book Value is the log of average total book value of firms headquartered in a MSA from 1996 to 2000 from the COMPUSTAT. Log Total Income is the log of average total income in a MSA from 1996 to 2000 from the Bureau of Economic Analysis. RESIDUAL BOOK VALUE is the residual of log total book value on the quadratic form of log total income using years from 1996 to 2000. Supply Elasticity is the housing supply elasticity of MSAs from Saiz (2010). Home Price Appreciation is the annual percentage increase in MSA home price index from the Federal Housing Finance Agency. % Change in Total MTG Origination is the annual percentage change in total mortgage origination amount in a MSA from the Home Mortgage Disclosure Act Data. % Change in Second Home MTG Origination is the annual percentage change in second home mortgage origination amount in a MSA. We report population-weighted mean and standard deviation using MSA population in 2000. Panel B reports the summary statistics for large MSAs with average population above 750000.

Panel A: All MSAs

Variable	Obs	Mean	Std. Dev.	Min	Max
log Total Book Value	277	23.50	2.94	10.37	27.36
log Total Income	277	17.81	1.50	14.33	20.29
RESIDUAL BOOK VALUE	277	0.06	1.40	-9.41	5.60
Supply Elasticity	277	1.72	1.05	0.63	12.15
From 2000 to 2006					
Annual Home Price Appreciation	277	0.09	0.05	0.01	0.18
Annual % Change in Total MTG Origination	277	0.27	0.09	0.08	1.01
Annual % Change in Second Home MTG Origination	277	0.36	0.25	0.05	5.45
From 2007 to 2010					
Annual Home Price Appreciation	277	-0.05	0.05	-0.22	0.03
Annual % Change in Total MTG Origination	277	-0.08	0.12	-0.42	0.30
Annual % Change in Second Home MTG Origination	277	-0.16	0.16	-0.44	2.39

Panel B: Large MSAs

Variable	Obs	Mean	Std. Dev.	Min	Max
log Total Book Value	65	24.79	1.64	18.40	27.36
log Total Income	65	18.51	1.03	16.38	20.29
RESIDUAL BOOK VALUE	65	0.09	0.96	-3.51	2.23
log RATIO	65	-0.59	1.00	-4.98	1.33
Supply Elasticity	65	1.44	0.79	0.63	4.00
From 2000 to 2006					
Annual Home Price Appreciation	65	0.10	0.05	0.03	0.18
Annual % Change in Total MTG Origination	65	0.28	0.09	0.13	0.58
Annual % Change in Second Home MTG Origination	65	0.37	0.14	0.18	0.90
From 2007 to 2010					
Annual Home Price Appreciation	65	-0.06	0.05	-0.21	0.01
Annual % Change in Total MTG Origination	65	-0.08	0.11	-0.40	0.13
Annual % Change in Second Home MTG Origination	65	-0.18	0.09	-0.42	0.08

Table 3: RESIDUAL BOOK VALUE

We report estimates of the relationship of log Total Income on log Total Book Value. Log Total Book Value is the log of average total book value of firms headquartered in a MSA from 1996 to 2000 and log Total Income is the log of average total income in a MSA from 1996 to 2000. Panel A reports regression results with various specifications. Column (1) shows the result with linear form, column (2) shows the result with quadratic form, and column (3) shows the result with cubic form. Panel B reports the correlation matrix of RESIDUAL BOOK VALUEs from different specifications. The table reports point estimates with t-statistics in parentheses. We use robust standard errors. ***, **, * denotes 1%, 5%, and 10% statistical significance.

Panel A: Various specifications for RESIDUAL BOOK VALUE

	(1)	(2)	(3)
Variables	log Total Book Value		
log Total Income	2.275***	13.52***	84.66*
	(17.99)	(4.86)	(1.88)
log Total Income Square		-0.338***	-4.572*
		(-4.05)	(-1.71)
log Total Income Cubic			0.084
			(1.58)
Constant	-16.39***	-109.2***	-505.9**
	(-8.06)	(-4.75)	(-2.01)
Observations	277	277	277
Adjusted R-squared	.539	0.563	0.566

Panel B: Correlation Matrix of RESIDUAL BOOK VALUEs

	Linear	Quadratic	Cubic
RESIDUAL BOOK VALUE: Linear Form	1		
RESIDUAL BOOK VALUE: Quadratic Form	0.971	1	
RESIDUAL BOOK VALUE: Cubic Form	0.967	0.996	1

Table 4: RESIDUAL BOOK VALUE and Sand States

We report estimates of the relationship of Sand State dummy on RESIDUAL BOOK VALUE. Columns (1)-(2) report the results on RESIDUAL BOOK VALUE. RESIDUAL BOOK VALUE is the residual of log total book value on the quadratic form of log total income using years from 1996 to 2000. Sand State dummy equals to 1 for MSAs in Arizona, California, Florida, or Nevada. Column (1) shows the result without division fixed effects and column (2) shows the result with division fixed effects. Columns (3)-(4) report the results on the rankings of RESIDUAL BOOK VALUE among 277 MSAs. The MSA with highest RESIDUAL BOOK VALUE gets 1st RANK and the MSA with lowest RESIDUAL BOOK VALUE gets 277th RANK. Column (3) shows the result without division fixed effects and column (4) shows the result with division fixed effects. We use MSA population in 2000 as a weight for all regression results. The table reports point estimates with t-statistics in parentheses. We use robust standard errors. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	(1)	(2)	(3)	(4)
Variables	RESIDUAL BOOK VALUE		RANK	
Sand State Dummy	-1.135***	-0.771***	63.98***	51.81***
	(-4.02)	(-3.46)	(4.78)	(4.09)
Constant	0.324**		131.8***	
	(2.03)		(13.39)	
Observations	277	277	277	277
R-squared	0.120	0.157	0.186	0.257
Division FE	No	Yes	No	Yes

Table 5: RESIDUAL BOOK VALUE and Mortgage Origination

We report estimates of the relationship of RESIDUAL BOOK VALUE on annual % change in mortgage origination. Panel A shows the results with all MSAs. Columns (1)-(2) report the results on annual % Change in Total Mortgage Origination. Column (1) shows result for the period of 2000 to 2006 and column (2) shows the result for the period of 2007 to 2010. We control for RESIDUAL BOOK VALUE, Supply Elasticity, High Income Growth Dummy, High POP Growth Dummy and include state fixed effects. RESIDUAL BOOK VALUE is the residual of log total book value on the quadratic form of log total income using years from 1996 to 2000. High Income Growth Dummy is an indicator for MSAs with above median income growth rate during 1990 to 2000. High POP Growth Dummy is an indicator for MSAs with above median population growth during 1990 to 2000. Columns (3)-(4) report the results on annual % Change in Second Home Mortgage Origination. Column (3) shows result for the period of 2000 to 2006 and column (4) shows the result for the period of 2007 to 2010. Panel B shows the results only with large MSAs whose average population is above 750000. We use MSA population in 2000 as a weight for all regression results. The table reports point estimates with t-statistics in parentheses. We use robust standard errors. ***, **, * denotes 1%, 5%, and 10% statistical significance.

Panel A: All MSAs

	Annual % Change in Total Mortgage Origination		Annual % Change in Second Home Mortgage Origination	
	2000-2006	2007-2010	2000-2006	2007-2010
Variables	(1)	(2)	(3)	(4)
RESIDUAL BOOK VALUE	-0.013***	0.018***	-0.033	0.008
	(-3.333)	(4.04)	(-1.51)	(1.46)
Supply Elasticity	-0.008*	0.009	-0.010	0.048**
	(-1.88)	(1.18)	(-0.68)	(2.10)
High Income Growth Dummy	-0.050***	0.056**	-0.033	-0.002
	(-3.54)	(2.48)	(-0.95)	(-0.07)
High POP Growth Dummy	0.070***	-0.072***	0.021	-0.054**
	(3.86)	(-2.77)	(0.48)	(-2.35)
Observations	277	277	277	277
R-squared	0.748	0.741	0.283	0.445
State FE	Yes	Yes	Yes	Yes

Panel B: Large MSAs

	Annual % Change in Total Mortgage Origination		Annual % Change in Second Home Mortgage Origination	
	2000-2006	2007-2010	2000-2006	2007-2010
Variables	(1)	(2)	(3)	(4)
RESIDUAL BOOK VALUE	-0.042***	0.058***	-0.032**	0.044***
	(-4.33)	(4.31)	(-2.20)	(3.34)
Supply Elasticity	0.008	0.021	0.042	0.028
	(0.42)	(0.84)	(0.93)	(0.83)
High Income Growth Dummy	-0.051***	0.078***	-0.081**	0.037
	(-3.09)	(3.28)	(-2.73)	(1.26)
High POP Growth Dummy	0.053*	-0.047	0.018	0.004
	(1.74)	(-1.25)	(0.37)	(0.09)
Observations	65	65	65	65
R-squared	0.968	0.955	0.965	0.938
State FE	Yes	Yes	Yes	Yes

Table 6: RESIDUAL BOOK VALUE and Home Price Appreciation

We report estimates of the relationship of RESIDUAL BOOK VALUE on annual home price appreciation. Panel A shows the results with all MSAs. Columns (1)-(2) show results for the period of 2000 to 2006 and columns (3)-(4) show the results for the period of 2007 to 2010. Columns (1) and (3) control for RESIDUAL BOOK VALUE, Supply Elasticity, High Income Growth Dummy, High POP Growth Dummy and state fixed effects. RESIDUAL BOOK VALUE is the residual of log total book value on the quadratic form of log total income using years from 1996 to 2000. High Income Growth Dummy is an indicator for MSAs with above median income growth rate during 1990 to 2000. High POP Growth Dummy is an indicator for MSAs with above median population growth during 1990 to 2000. Columns (2) and (4) additionally control for quintile dummies of Supply Elasticity interacted with RESIDUAL BOOK VALUE. For brevity, coefficients on quintile dummies of Supply Elasticity are not reported. Panel B shows the results only with large MSAs whose average population is above 750000. We use MSA population in 2000 as a weight for all regression results. The table reports point estimates with t-statistics in parentheses. We use robust standard errors. ***, **, * denotes 1%, 5%, and 10% statistical significance.

Panel A: All MSAs	Annual Home Price Appreciation			
	2000 - 2006		2007 - 2010	
Variables	(1)	(2)	(3)	(4)
RESIDUAL BOOK VALUE	-0.002**	-0.008**	0.003**	0.008**
	(-2.24)	(-2.39)	(2.51)	(2.24)
Supply Elasticity	-0.006***	-0.007*	0.006***	-0.001
	(-3.55)	(-1.73)	(3.05)	(-0.37)
High Income Growth Dummy	-0.019***	-0.016**	0.014**	0.011**
	(-2.81)	(-2.45)	(2.31)	(2.12)
High POP Growth Dummy	0.010	0.008	-0.028***	-0.028***
	(1.53)	(1.25)	(-4.06)	(-4.75)
RESIDUAL BOOK VALUE × Elasticity Dummy (2/5)		0.004		-0.004
		(1.49)		(-1.08)
RESIDUAL BOOK VALUE × Elasticity Dummy (3/5)		0.006**		-0.004
		(2.09)		(-0.87)
RESIDUAL BOOK VALUE × Elasticity Dummy (4/5)		0.007*		-0.011**
		(1.95)		(-2.34)
RESIDUAL BOOK VALUE × Elasticity Dummy (5/5)		0.008**		-0.008*
		(2.44)		(-1.92)
Observations	277	277	277	277
R-squared	0.921	0.926	0.890	0.899
State FE	Yes	Yes	Yes	Yes

Panel B: Large MSAs	Annual Home Price Appreciation			
	2000 - 2006		2007 - 2010	
Variables	(1)	(2)	(3)	(4)
RESIDUAL BOOK VALUE	-0.010*	-0.032**	0.014***	0.016
	(-1.88)	(-2.72)	(3.35)	(1.08)
Supply Elasticity	-0.003	0.002	0.012	-0.009
	(-0.27)	(0.06)	(1.32)	(-0.28)
High Income Growth Dummy	-0.0245	-0.035**	0.016*	0.017*
	(-1.37)	(-2.70)	(1.73)	(2.05)
High POP Growth Dummy	-0.001	0.016	-0.021*	-0.022
	(-0.09)	(1.00)	(-1.74)	(-1.02)
RESIDUAL BOOK VALUE x Supply Elasticity Dummy (2/5)		0.025		0.002
		(1.32)		(0.06)
RESIDUAL BOOK VALUE x Supply Elasticity Dummy (3/5)		0.010		0.004
		(0.46)		(0.18)
RESIDUAL BOOK VALUE x Supply Elasticity Dummy (4/5)		0.024*		0.001
		(1.84)		(0.03)
RESIDUAL BOOK VALUE x Supply Elasticity Dummy (5/5)		0.030		-0.032
		(1.31)		(-1.23)
Observations	65	65	65	65
R-squared	0.960	0.985	0.966	0.972
State FE	Yes	Yes	Yes	Yes

Table 7: Explaining Sand States with RESIDUAL BOOK VALUE

We report estimates of the relationship of Sand State dummy on annual % change in mortgage origination and home price appreciation. Panel A shows the results with all MSAs. Columns (1)-(4) report the results on annual % Change in Total Mortgage Origination where (1)-(2) are for the period of 2000 to 2006 and (3)-(4) are for the period of 2007 to 2010. Columns (1) and (3) report the baseline regression with Sand State Dummy, Supply Elasticity, High Income Growth Dummy, High POP Growth Dummy and the quintile dummies of Supply Elasticity. Sand State dummy equals to 1 for MSAs in Arizona, California, Florida, or Nevada. High Income Growth Dummy is an indicator for MSAs with above median income growth rate during 1990 to 2000. High POP Growth Dummy is an indicator for MSAs with above median population growth during 1990 to 2000. For brevity, the coefficients on quintile dummies of Supply Elasticity are not reported. Columns (2) and (4) report the prediction results if RESIDUAL BOOK VALUE, the residual of log total book value on the quadratic form of log total income using years from 1996 to 2000, is zero for all MSAs. We first regress annual % Change in Total Mortgage Origination on Sand State Dummy, Supply Elasticity, quintile dummies of Supply Elasticity, RESIDUAL BOOK VALUE, High Income Growth Dummy, High POP Growth Dummy with state fixed effects. We compute the predicted value from the first-stage regression by setting RESIDUAL BOOK VALUE equals to zero. Then we use the predicted value as dependent variable for prediction results. First-stage regression results on mortgage originations are reported in Appendix Table 2. We report the percentage difference in baseline results and prediction results. Columns (5)-(8) report the results on annual % Change in Second Home Mortgage Origination. Columns (9)-(12) report the results on annual Home Price Appreciation. First-stage regression results on home price appreciation are reported in Table 5. Panel B shows the results only with large MSAs whose average population is above 750000. We use MSA population in 2000 as a weight for all regression results. The table reports point estimates with t-statistics in parentheses. We use robust standard errors. ***, **, * denotes 1%, 5%, and 10% statistical significance.

Panel A: All MSAs

	% Change in Total MTG Origination				% Change in Second Home MTG Origination				Home Price Appreciation			
	2000 - 2006		2007 - 2010		2000 - 2006		2007 - 2010		2000 - 2006		2007 - 2010	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Variables	Baseline	Prediction	Baseline	Prediction	Baseline	Prediction	Baseline	Prediction	Baseline	Prediction	Baseline	Prediction
Sand State Dummy	0.076***	0.042***	-0.085**	-0.043***	0.056***	0.013	0.007	0.036**	0.060***	0.051***	-0.075***	-0.065***
	(4.99)	(3.96)	(-2.27)	(-3.11)	(2.84)	(0.60)	(0.22)	(2.11)	(6.74)	(12.74)	(-8.06)	(-17.36)
Difference		*-45%*		*49%*		*-78%*		*377%*		*-15%*		*13%*
Supply Elasticity	-0.012	-0.016	-0.030	-0.025*	-0.010	-0.015	0.029	0.032*	-0.010	-0.011**	-0.009	-0.008**
	(-1.07)	(-1.38)	(-1.45)	(-1.66)	(-0.38)	(-0.64)	(0.46)	(1.73)	(-1.56)	(-2.50)	(-1.39)	(-2.00)
High Income Growth Dummy	-0.042**	-0.032***	0.065**	0.053***	-0.048*	-0.035*	0.020	0.012	-0.021***	-0.018***	0.007	0.004
	(-2.37)	(-3.16)	(2.13)	(3.97)	(-1.97)	(-1.74)	(0.86)	(0.74)	(-2.90)	(-4.78)	(0.98)	(1.17)
High POP Growth Dummy	0.045**	0.032***	-0.123***	-0.106***	0.039	0.021	-0.084***	-0.072***	0.014**	0.011**	-0.022***	-0.018***
	(2.47)	(2.90)	(-4.28)	(-7.38)	(1.19)	(0.98)	(-3.01)	(-4.12)	(2.38)	(2.59)	(-3.04)	(-4.66)
Observations	277	277	277	277	277	277	277	277	277	277	277	277
R-squared	0.409	0.375	0.338	0.289	0.086	0.205	0.107	0.193	0.708	0.697	0.731	0.757

Panel B: Large MSAs

	% Change in Total MTG Origination				% Change in Second Home MTG Origination				Home Price Appreciation			
	2000 - 2006		2007 - 2010		2000 - 2006		2007 - 2010		2000 - 2006		2007 - 2010	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Variables	Baseline	Prediction	Baseline	Prediction	Baseline	Prediction	Baseline	Prediction	Baseline	Prediction	Baseline	Prediction
Sand State Dummy	0.068***	-0.013	-0.063	0.019	0.059**	0.0001	-0.004	0.042	0.052***	0.009	-0.066***	-0.044***
	(3.99)	(-0.67)	(-1.30)	(0.36)	(2.33)	(0.01)	(-0.10)	(1.13)	(6.06)	(0.62)	(-6.47)	(-4.29)
Difference		*-119%*		*130%*		*-100%*		*1150%*		*-84%*		*33%*
Supply Elasticity	0.034	0.033	0.021	0.021	0.104	0.103	0.056	0.057	0.011	0.011	0.021	0.021
	(0.88)	(0.76)	(0.33)	(0.31)	(1.13)	(1.12)	(1.38)	(1.43)	(0.74)	(0.53)	(1.50)	(1.51)
High Income Growth Dummy	-0.056***	-0.050***	0.061	0.056	-0.059*	-0.055*	0.005	0.002	-0.032***	-0.029*	0.005	0.003
	(-2.82)	(-2.70)	(1.44)	(1.09)	(-1.89)	(-1.78)	(0.15)	(0.04)	(-4.03)	(-1.89)	(0.51)	(0.34)
High POP Growth Dummy	0.057**	0.063**	-0.130***	-0.137***	0.064	0.069	-0.027	-0.031	0.030**	0.034**	-0.020*	-0.022**
	(2.26)	(2.50)	(-2.78)	(-2.89)	(1.46)	(1.61)	(-0.68)	(-0.79)	(2.62)	(2.19)	(-1.87)	(-2.19)
Observations	65	65	65	65	65	65	65	65	65	65	65	65
R-squared	0.479	0.383	0.327	0.202	0.287	0.246	0.183	0.128	0.728	0.580	0.750	0.689

Table 8: RATIO on Mortgage Origination and Home Price Appreciation

We report estimates of the relationship of log RATIO on annual % change in mortgage origination and home price appreciation. We use large MSAs whose average population is above 750000. Columns (1)-(2) report the results on annual % Change in Total Mortgage Origination, column (3)-(4) report the results on annual % Change in Second Home Mortgage Origination, and columns (5)-(6) report the results on annual home price appreciation. Independent variables are log RATIO, Supply Elasticity, High Income Growth Dummy and High POP Growth Dummy. log RATIO is the log ratio of the total book value of firms headquartered in a MSA to the income in that MSA as in Hong, Kubik, and Stein (2008). Due to the MSAs with RATIO equal to zero, we add 0.00001 to RATIO before we take log. High Income Growth Dummy is an indicator for MSAs with above median income growth rate during 1990 to 2000. High POP Growth Dummy is an indicator for MSAs with above median population growth during 1990 to 2000. We include state fixed effects. We use MSA population in 2000 as a weight for all regression results. The table reports point estimates with t-statistics in parentheses. We use robust standard errors. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	% Change in Total MTG Origination		% Change in Second Home MTG Origination		Home Price Appreciation	
	2000 - 2006	2007 - 2010	2000 - 2006	2007 - 2010	2000 - 2006	2007 - 2010
Variables	(1)	(2)	(3)	(4)	(5)	(6)
log RATIO	-0.0381***	0.0447**	-0.0352**	0.0331**	-0.00885*	0.0105*
	(-3.854)	(2.719)	(-2.688)	(2.351)	(-1.939)	(1.947)
Supply Elasticity	-0.0110	0.0412	0.0224	0.0424	-0.00716	0.0163
	(-0.525)	(1.564)	(0.487)	(1.188)	(-0.684)	(1.624)
High Income Growth Dummy	-0.0348*	0.0547	-0.0596*	0.0211	-0.0211	0.0110
	(-1.992)	(1.389)	(-1.994)	(0.521)	(-1.187)	(0.807)
High POP Growth Dummy	0.0596*	-0.0462	0.0205	0.00391	0.000169	-0.0209
	(2.045)	(-1.132)	(0.432)	(0.0815)	(0.0111)	(-1.693)
Observations	65	65	65	65	65	65
R-squared	0.967	0.941	0.970	0.923	0.960	0.959
State FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 9: Summary Statistics

We report summary statistics of the Survey of Consumer Finance for 1995, 1998, 2001, 2004, 2007, and 2010 waves. The sample consists of households in MSAs with population greater than 750,000. HAVE INVESTMENT REAL ESTATE is an indicator that the family owns investment real estate. HAVE DIRECTLY-HELD STOCKS is an indicator that the family directly owns stock. % INVESTMENT REAL ESTATE is the share of a family's assets invested in investment real estate. % DIRECTLY-HELD STOCKS is the share a family's assets invested directly in stock. log RATIO is the log ratio of the total book value of firms headquartered in a MSA to the income in that MSA as in Hong, Kubik, and Stein (2008). Due to the MSAs with RATIO equal to zero, we add 0.00001 to RATIO before we take log. log HOUSE PRICE ZIP CODE is a home price index at zip code level from Zillow. FAMILY SIZE is the number of people in each Primary Economic Unit. log HOUSEHOLD INCOME is the log of household income. HOUSEHOLD NET WORTH is the household's net worth. UNATTACHED FEMALE is an indicator that the family is headed by an unattached woman. AGE, RACE, EDUCATION, and FAMILY STRUCTURE are also reported. LWP stands for the "living with partner" in the FAMILY STRUCTURE variable.

Variables	Obs	Mean	Std.	Max	Min
HAVE INVESTMENT REAL ESTATE	89956	0.13	0.34	1	0
HAVE VACATION HOME	89956	0.06	0.24	1	0
HAVE DIRECTLY-HELD STOCKS	89956	0.21	0.41	1	0
% INVESTMENT REAL ESTATE IN TOTAL ASSETS	89956	0.03	0.10	1	0
% VACATION HOME IN TOTAL ASSETS	89956	0.01	0.04	0.87	0
% DIRECTLY-HELD STOCKS IN TOTAL ASSETS	87700	0.02	0.08	1	0
log RATIO	89956	-1.13	1.08	1.11	-11.51
MSA Unemployment Rate	89956	5.65	2.26	15.87	2.14
MSA Home Price Index	89956	162.67	53.67	323.15	100.75
MSA Housing Affordability Index	86411	119.75	57.82	383.40	51.17
log HOUSE PRICE ZIP CODE	58964	12.22	0.64	15.14	10.40
FAMILY SIZE	89956	2.44	1.42	13	1
log HOUSEHOLD INCOME	89956	10.87	1.13	19.64	0
HOUSEHOLD NET WORTH	89956	591580	4052609	4.80E+09	-2.74E+07
UNATTACHED FEMALE	89956	0.27	0.45	1	0
AGE					
<35	89956	0.23	0.42	1	0
35 – 49	89956	0.33	0.47	1	0
50 – 64	89956	0.24	0.43	1	0
≥65	89956	0.19	0.40	1	0
RACE					
White	89956	0.68	0.47	1	0
Black	89956	0.16	0.36	1	0
Hispanic	89956	0.10	0.31	1	0
Asian & Other	89956	0.05	0.21	1	0
EDUCATION					
Less than High School	89956	0.13	0.34	1	0
High School or GED only	89956	0.28	0.45	1	0
Some college (or Associate's), but no Bachelor's	89956	0.25	0.43	1	0
Bachelor's	89956	0.21	0.41	1	0
MA or MS(non-MBA), plus nursing degree	89956	0.07	0.26	1	0
Advanced (PhD, JD, MBA, MD)	89956	0.06	0.24	1	0
FAMILY STRUCTURE					
Not Married/LWP+Children	89956	0.13	0.33	1	0
Not Married/LWP+no Children+Head(<55)	89956	0.16	0.37	1	0
Not Married/LWP+no Children+Head(≥55)	89956	0.13	0.34	1	0
Married/LWP+Children	89956	0.32	0.47	1	0
Married/LWP+no Children	89956	0.26	0.44	1	0

Table 10: Distribution of Average FICO Score at Zip Code Level

The entries are measures of the distribution of average FICO scores of households aggregated to the zip code level. The sample includes all zip codes in the US. The zip codes are sorted into quartiles based on average FICO score.

	Mean	Min	Max
Lowest FICO Quartile	621	439	643
Second FICO Quartile	663	644	680
Third FICO Quartile	695	681	709
Highest FICO Quartile	724	710	822

Table 11: Investment Real Estate Ownership, Directly-held Stocks Ownership and log RATIO

We report logit estimates of the relationship between demographic characteristics including log RATIO and the likelihood that a household owns investment real estate and stocks. The dependent variable in columns (1) through (2) is an indicator that the family owns investment real estate. The dependent variable in columns (3) through (4) is an indicator that the family directly owns stock. The independent variables include log RATIO, the log ratio of the total book value of firms headquartered in a MSA to the income in that MSA as in Hong, Kubik, and Stein (2008), log HOUSEHOLD INCOME, the usual income of the family and log HOUSEHOLD NET WORTH, the wealth of family (not including real estate in columns (1) and (2) and not including stock holdings in columns (3) and (4)). Due to the MSAs with RATIO equal to zero, we add 0.00001 to RATIO before we take log. High FICO Zip Code is a dummy for the household living in a zip code in the top quartile of the average FICO score distribution. UNATTACHED FEMALE is an indicator that the family is headed by an unattached woman. Other Household Controls include family structure and a linear and square term of the number of people in the family. The table reports point estimates with t-statistics in parentheses. All the standard errors are clustered at the MSA level. ***, **, * denotes 1%, 5%, and 10% statistical significance. The average marginal effect is in brackets.

Variables		(1) HAVE INVESTMENT REAL ESTATE	(2)	(3) HAVE DIRECTLY-HELD STOCK	(4)
log RATIO		-0.0751*	-0.0695**	0.126***	0.13***
		(-1.91)	(-2.37)	(4.01)	(3.71)
		[-.0074]	[-.0068]	[.0164]	[.0169]
High FICO Zip Code		-0.309***	-0.222***	0.21***	0.192***
		(-4.11)	(-3.46)	(3.3)	(3.14)
log HOUSEHOLD INCOME		0.218***	0.218***	0.348***	0.351***
		(4.63)	(4.83)	(6.49)	(6.44)
log HOUSEHOLD NET WORTH		0.594***	0.598***	0.428***	0.429***
		(14.45)	(14.59)	(13.05)	(13.58)
UNATTACHED FEMALE		-0.269**	-0.27**	-0.219***	-0.222***
		(-2.24)	(-2.18)	(-3.15)	(-3.15)
Age	Age 35-49	0.288**	0.29**	-0.231**	-0.227**
		(2.38)	(2.44)	(-2.18)	(-2.10)
	Age 50-64	0.546***	0.565***	-0.349***	-0.336***
		(4.4)	(4.67)	(-3.20)	(-3.08)
	Age 65+	0.498***	0.513***	-0.098	-0.09
		(3.66)	(3.95)	(-0.90)	(-0.84)
Race	Black	0.344***	0.378***	-0.496***	-0.501***
		(3.19)	(3.67)	(-7.33)	(-7.29)
	Hispanic	0.199	0.179	-0.975***	-0.957***
		(1.36)	(1.11)	(-5.98)	(-5.94)
	Asian and Other	0.116	0.147	0.084	0.089
		(0.89)	(1.12)	(0.89)	(0.92)
Education	High School Degree	-0.124	-0.139	0.474***	0.481***
		(-0.86)	(-0.15)	(3.49)	(3.56)
	Some College	0.01	-0.007	0.799***	0.824***
		(0.12)	(-0.13)	(5.19)	(5.35)
	Bachelors Degree	-0.059	-0.076	1.161***	1.185***
		(-0.42)	(-0.56)	(7.79)	(7.8)
	MA or MS	0.027	0.009	1.151***	1.173***
		(0.15)	(0.16)	(6.85)	(6.98)
	Advanced Degree	-0.197	-0.225	1.032***	1.059***
		(-1.34)	(-1.51)	(6.29)	(6.3)
Other Household Controls		Yes	Yes	Yes	Yes
Year Effects		Yes	—	Yes	—
Year × Census Division Effects		No	Yes	No	Yes
Observations		89956	89956	89956	89956

Table 12: % Share of Investment Property, % Share of Directly-held Stocks and log RATIO

We report tobit estimates of the relationship between log RATIO and the amount that a family owns of investment real estate and stocks. The dependent variable in columns (1) and (2) is the share of a family's assets invested in investment real estate. The dependent variable in columns (3) and (4) is the share the family's assets invested directly in stock. The independent variables include log RATIO, the log ratio of the total book value of firms headquartered in a MSA to the income in that MSA as in Hong, Kubik, and Stein (2008), log HOUSEHOLD INCOME, the usual income of the family and log HOUSEHOLD NET WORTH, the wealth of family (not including real estate in columns (1) and (2) and not including stock holdings in columns (3) and (4)). Due to the MSAs with RATIO equal to zero, we add 0.00001 to RATIO before we take log. High FICO Zip Code is a dummy for the household living in a zip code in the top quartile of the average FICO score distribution. UNATTACHED FEMALE is an indicator that the family is headed by an unattached woman. Other Household Controls include family structure and a linear and square term of the number of people in the family. The table reports point estimates with t-statistics in parentheses. All the standard errors are clustered at the MSA level. ***, **, * denotes 1%, 5%, and 10% statistical significance.

		(1)	(2)	(3)	(4)
Variables		% INVESTMENT REAL ESTATE		% DIRECTLY-HELD STOCKS	
log RATIO		-0.0144**	-0.0111**	0.015***	0.0144***
		(-2.11)	(-1.99)	(4.5)	(4.04)
High FICO Zip Code		-0.0551***	-0.043***	0.0254***	0.0246***
		(-3.96)	(-3.21)	(3.22)	(3.42)
log HOUSEHOLD INCOME		0.0406***	0.0388***	0.0439***	0.044***
		(3.87)	(3.77)	(5.69)	(5.79)
log HOUSEHOLD NET WORTH		0.0959***	0.0956***	0.0269***	0.0268***
		(12.14)	(12.45)	(5.69)	(5.94)
UNATTACHED FEMALE		-0.0593**	-0.0575**	-0.0305***	-0.0313***
		(-2.31)	(-2.25)	(-3.68)	(-3.78)
Age	Age 35-49	0.0484**	0.0503**	-0.026***	-0.0259***
		(2.00)	(2.15)	(-2.79)	(-2.78)
	Age 50-64	0.098***	0.102***	-0.0305***	-0.0297***
		(4.15)	(4.45)	(-2.65)	(-2.58)
	Age 65+	0.0929***	0.0956***	0.0237**	0.0247**
		(3.32)	(3.51)	(2.24)	(2.38)
Race	Black	0.079***	0.0843***	-0.0639***	-0.0638***
		(3.45)	(3.81)	(-8.08)	(-8.23)
	Hispanic	0.0427	0.0343	-0.104***	-0.104***
		(1.47)	(1.06)	(-5.91)	(-6.15)
	Asian and Other	0.0337	0.0346	-0.0012	-0.0001
		(1.21)	(1.21)	(-0.11)	(-0.01)
Education	High School Degree	-0.0387	-0.0399	0.0657***	0.0648***
		(-1.28)	(-1.31)	(4.35)	(4.29)
	Some College	-0.0075	-0.0095	0.102***	0.103***
		(-0.30)	(-0.37)	(5.96)	(6.09)
	Bachelors Degree	-0.0102	-0.0121	0.14***	0.14***
		(-0.35)	(-0.41)	(8.28)	(8.19)
	MA or MS	0.0073	0.0056	0.138***	0.139***
		(0.24)	(0.19)	(7.8)	(7.99)
	Advanced Degree	-0.0277	-0.03	0.133***	0.133***
		(-0.90)	(-0.96)	(7.35)	(7.19)
Other Household Controls		Yes	Yes	Yes	Yes
Year Effects		Yes	—	Yes	—
Year × Census Division Effects		No	Yes	No	Yes
Observations		89956	89956	87700	87700

Table 13: The Difference in Investment Home Ownership Between High and Low Credit Risk Households Within a MSA by log RATIO

We report estimates of the relationship of log RATIO on a family's holdings of investment real estate, splitting up the effect of log RATIO for households living in low and high credit risk zip codes. The sample consists of families who live in a zip code in the highest and lowest 25% of the credit risk distribution (measured by average FICO scores). The dependent variable in columns (1) and (2) is an indicator for whether the family owns investment real estate. The dependent variable in columns (3) and (4) is the share of investment real estate of a familys assets. Columns (1) and (3) replicate the results in Table 4 and 5 to show that main results are still valid in the subsample. The independent variables are the same as in Table 4 except for a couple of differences. log RATIO x High FICO Zip Code is the interaction of that indicator and log RATIO. Year × MSA Effects are a full set of MSA by year dummy interactions. Because of the inclusion of Year × MSA Effects, the coefficient on log RATIO is not identified. The coefficients for the first two columns are from a OLS model. The coefficients for the last two columns are from a tobit model. The table reports point estimates with t-statistics in parentheses. All the standard errors are clustered at the MSA level. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	(1)	(2)	(3)	(4)
Variables	HAVE INVESTMENT REAL ESTATE		% INVESTMENT REAL ESTATE	
log RATIO × High FICO Zip Code		-0.0218**		-0.0408***
		(-2.14)		(-14.57)
log RATIO	-0.0087	—	-0.0111	—
	(-1.38)		(-1.33)	
High FICO Zip Code	-0.0209	-0.0334	-0.0348	-0.0444***
	(-1.38)	(-1.61)	(-1.63)	(-11.24)
log HOUSEHOLD INCOME	0.0448***	0.044***	-0.0022	0.0052***
	(7.63)	(7.47)	(-0.24)	(13.8)
log HOUSEHOLD NET WORTH	0.0146***	0.014***	0.0372***	0.0101***
	(7.09)	(6.22)	(4.28)	(30.7)
UNATTACHED FEMALE	-0.0142	-0.0154	-0.015	0.0061
	(-0.89)	(-0.98)	(-0.68)	(1.19)
Other Household Controls	Yes	Yes	Yes	Yes
Year Effects	Yes	—	Yes	—
Year × MSA Effects	No	Yes	No	Yes
Observations	41845	41845	41845	41845

Figure 1: RESIDUAL BOOK VALUE and Sand States

The figure shows the scatter plot of MSA log Total Income and MSA log Total Book value with quadratic fitted line. Log Total Book Value is the log of total book value of firms headquartered in a MSA and log Total Income is the log of total income in a MSA. We average log Total Income and log Total Book value from 1996 to 2000 for the plot. RESIDUAL BOOK VALUE, which is the residual of log Total Book Value on the quadratic form of log Total Income using years from 1996 to 2000, is the orthogonal distance to the fitted line. We report the names of MSAs in sand states (AZ, CA, FL, NV).

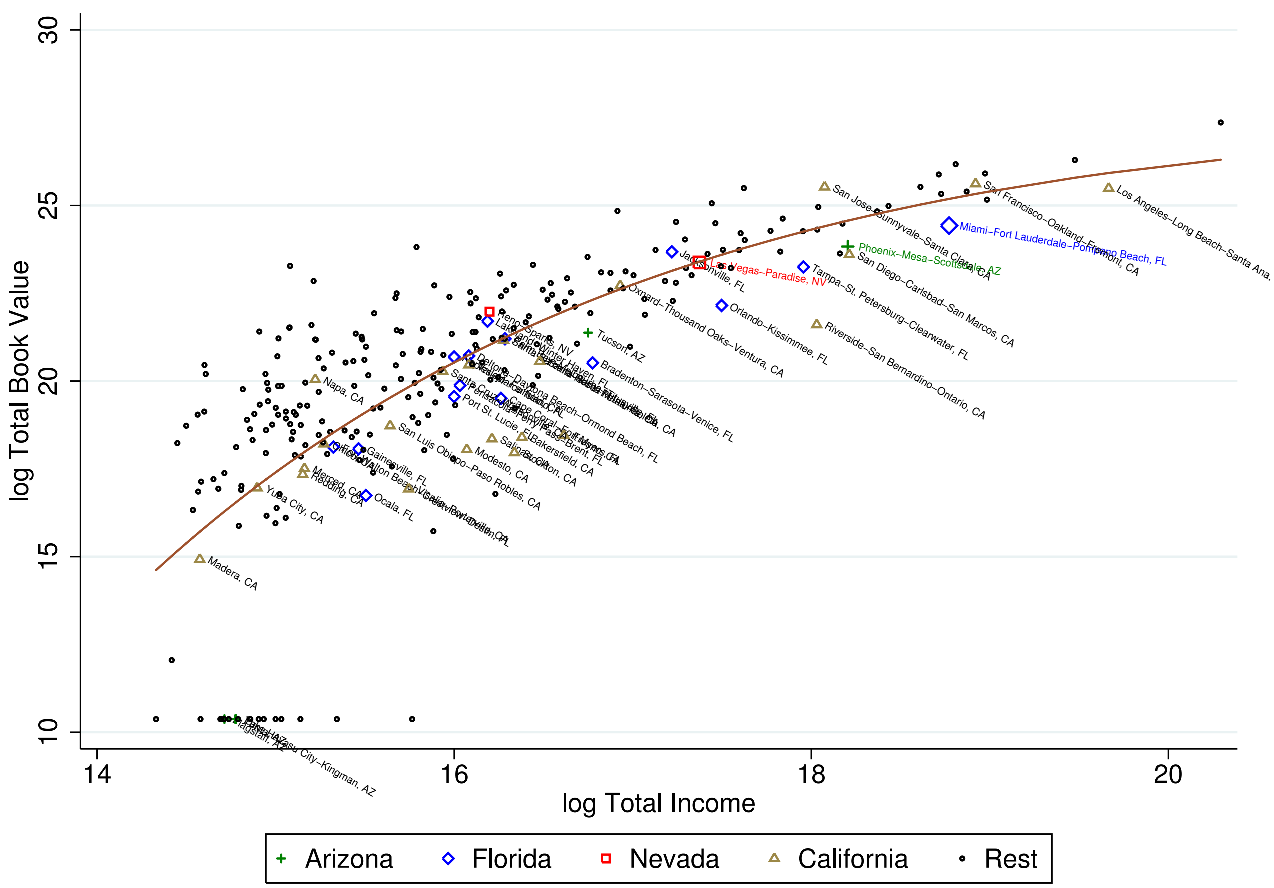
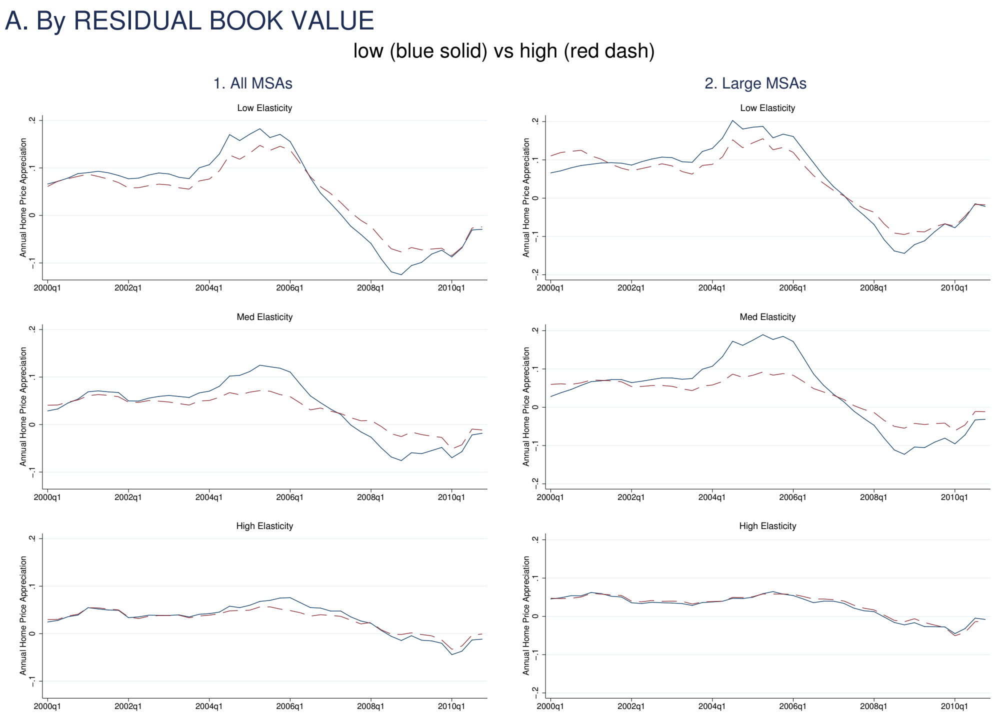


Figure 2: Home Price Appreciation by Elasticity and by Book Value

The figure shows the time-series of average home price appreciation by the groups of elasticity and book value. Panel A independently sorts MSAs into six groups: 3 groups of supply elasticity and 2 groups of RESIDUAL BOOK VALUE. RESIDUAL BOOK VALUE is the residual of log total book value on the quadratic form of log total income using years from 1996 to 2000. First column shows the graph using all MSAs and second column shows the graph using only large MSAs with average population more than 750000. Blue solid line indicates low RESIDUAL BOOK VALUE group and red dash line indicates high RESIDUAL BOOK VALUE group. Panel B independently sorts MSAs into six groups: 3 groups of supply elasticity and 2 groups of RATIO. RATIO is the ratio of the total book value of firms headquartered in a MSA to the income in that MSA as in Hong, Kubik, and Stein (2008). First column shows the graph using all MSAs and second column shows the graph using only large MSAs. Blue solid line indicates low RATIO group and red dash line indicates high RATIO group.



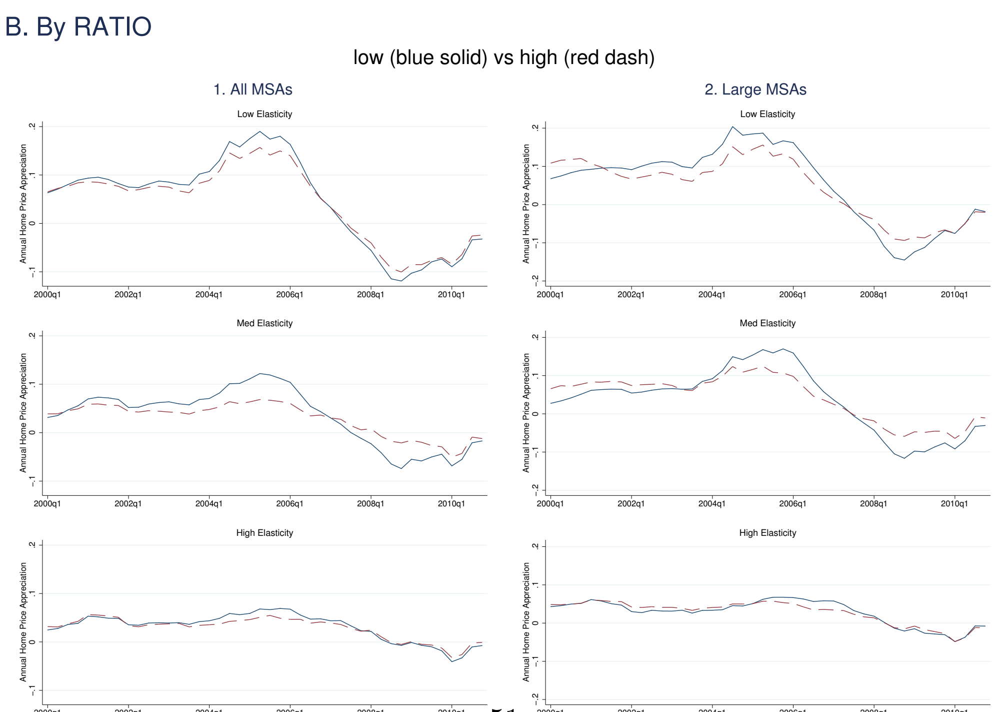


Figure 3: Geographic Distribution of RATIO Variable across the US

The figure shows the average RATIO variable for MSAs in the US with population greater than 750,000 between 1996 and 2010. RATIO is the ratio of the total book value of firms headquartered in a MSA to the income in that MSA as in Hong, Kubik, and Stein (2008).

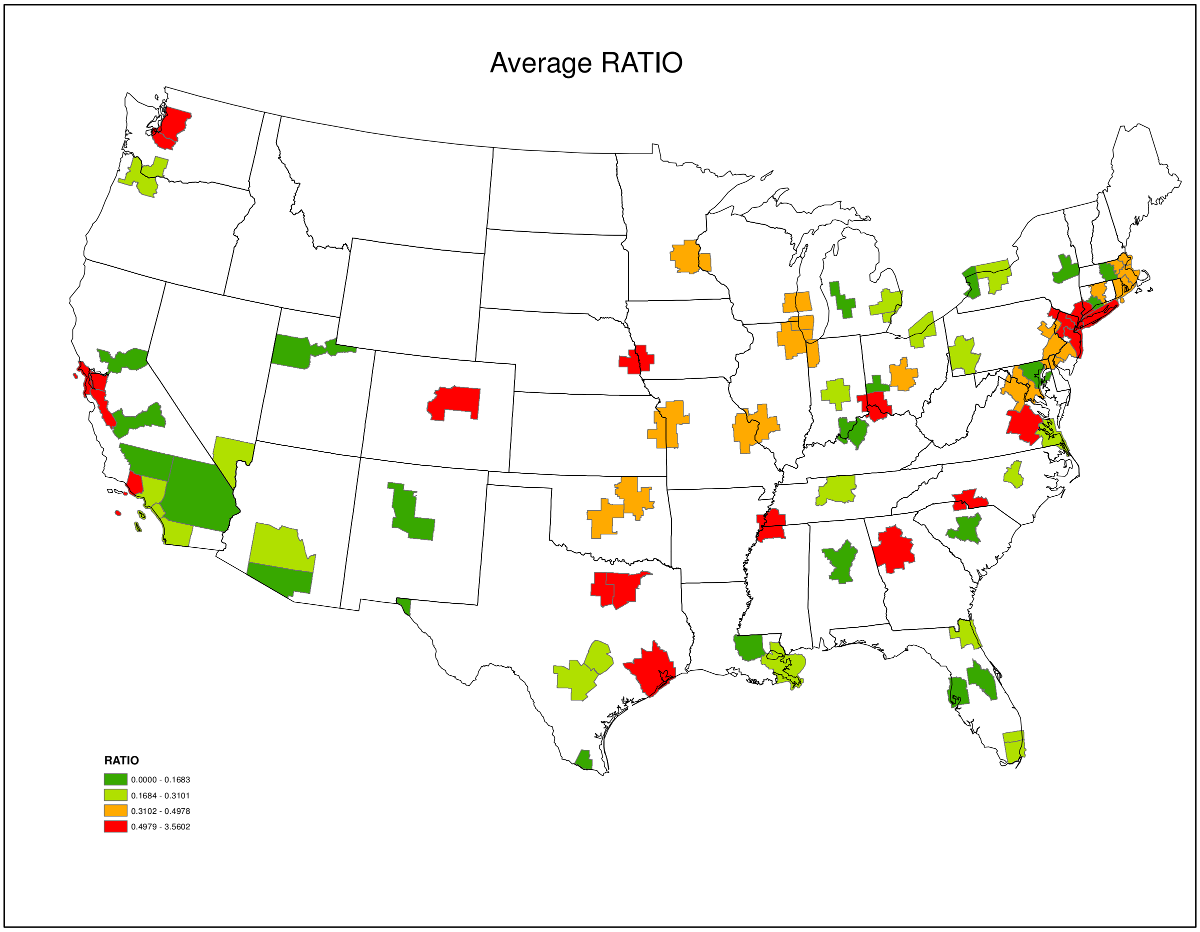


Figure 4: MSA/Year Residualized Scatter Plot

The figure shows scatter plots of the residualized MSA/YEAR log RATIO with the residualized MSA/YEAR investment real estate ownership (HAVE INVESTMENT REAL ESTATE) or the residualized MSA/YEAR % share of investment real estate in households' asset (% INVESTMENT REAL ESTATE). Panel A shows the scatter plot of residualized investment real estate ownership. Panel B shows the scatter plot of residualized % share of investment real estate in households' assets. Log RATIO is the log of ratio of the total book value of firms headquartered in a MSA to the income in that MSA as in Hong, Kubik, and Stein (2008). Due to the MSAs with RATIO equal to zero, we add 0.00001 to RATIO before we take log.

Table A1: List of Large MSAs

We report the list of large MSAs with more than 750000 population on average during 2000-2010. MSAs are sorted and ranked by RESIDUAL BOOK VALUE. RESIDUAL BOOK VALUE is the residual of log total book value on the quadratic form of log total income using years from 1996 to 2000. MSAs in sand states (AZ, CA, FL, NV) are reported in bold.

Code	MSA name	RESIDUAL BOOK VALUE	Rank
36540	Omaha-Council Bluffs, NE-IA	2.23	1
35620	New York-Northern New Jersey-Long Island, NY-NJ-PA	1.62	2
14860	Bridgeport-Stamford-Norwalk, CT	1.59	3
16740	Charlotte-Gastonia-Concord, NC-SC	1.45	4
40060	Richmond, VA	1.28	5
41940	**San Jose-Sunnyvale-Santa Clara, CA**	0.97	6
19100	Dallas-Fort Worth-Arlington, TX	0.86	7
25540	Hartford-West Hartford-East Hartford, CT	0.85	8
13820	Birmingham-Hoover, AL	0.70	9
32820	Memphis, TN-MS-AR	0.69	10
10900	Allentown-Bethlehem-Easton, PA-NJ	0.68	11
26420	Houston-Sugar Land-Baytown, TX	0.64	12
16980	Chicago-Naperville-Joliet, IL-IN-WI	0.60	13
46140	Tulsa, OK	0.55	14
47900	Washington-Arlington-Alexandria, DC-VA-MD-WV	0.47	15
27260	**Jacksonville, FL**	0.47	16
19740	Denver-Aurora, CO	0.44	17
39580	Raleigh-Cary, NC	0.42	18
17140	Cincinnati-Middletown, OH-KY-IN	0.39	19
12060	Atlanta-Sandy Springs-Marietta, GA	0.38	20
18140	Columbus, OH	0.33	21
41860	**San Francisco-Oakland-Fremont, CA**	0.22	22
41620	Salt Lake City, UT	0.22	23
12940	Baton Rouge, LA	0.17	24
28140	Kansas City, MO-KS	0.14	25
33340	Milwaukee-Waukesha-West Allis, WI	0.10	26
19820	Detroit-Warren-Livonia, MI	0.08	27
37100	**Oxnard-Thousand Oaks-Ventura, CA**	0.08	28
34980	Nashville-Davidson–Murfreesboro–Franklin, TN	0.05	29
49340	Worcester, MA	0.05	30
14460	Boston-Cambridge-Quincy, MA-NH	0.03	31
41700	San Antonio, TX	0.02	32
42660	Seattle-Tacoma-Bellevue, WA	0.01	33
19380	Dayton, OH	-0.04	34
33460	Minneapolis-St. Paul-Bloomington, MN-WI	-0.07	35
29820	**Las Vegas-Paradise, NV**	-0.11	36
17460	Cleveland-Elyria-Mentor, OH	-0.13	37
26900	Indianapolis-Carmel, IN	-0.13	38
35380	New Orleans-Metairie-Kenner, LA	-0.14	39
38300	Pittsburgh, PA	-0.19	40
41180	St. Louis, MO-IL	-0.20	41
31100	**Los Angeles-Long Beach-Santa Ana, CA**	-0.25	42
37980	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	-0.28	43
40380	Rochester, NY	-0.29	44
24340	Grand Rapids-Wyoming, MI	-0.36	45
12420	Austin-Round Rock, TX	-0.40	46
47260	Virginia Beach-Norfolk-Newport News, VA-NC	-0.43	47
31140	Louisville/Jefferson County, KY-IN	-0.45	48
38900	Portland-Vancouver-Beaverton, OR-WA	-0.53	49
39300	Providence-New Bedford-Fall River, RI-MA	-0.57	50
36420	Oklahoma City, OK	-0.58	51
33100	**Miami-Fort Lauderdale-Pompano Beach, FL**	-0.86	52
46060	**Tucson, AZ**	-0.89	53
38060	**Phoenix-Mesa-Scottsdale, AZ**	-0.89	54
15380	Buffalo-Niagara Falls, NY	-0.94	55
12580	Baltimore-Towson, MD	-1.03	56
35300	New Haven-Milford, CT	-1.03	57
10740	Albuquerque, NM	-1.06	58
41740	**San Diego-Carlsbad-San Marcos, CA**	-1.13	59
45300	**Tampa-St. Petersburg-Clearwater, FL**	-1.15	60
36740	**Orlando-Kissimmee, FL**	-1.55	61
10580	Albany-Schenectady-Troy, NY	-1.78	62
40140	**Riverside-San Bernardino-Ontario, CA**	-2.89	63
12540	**Bakersfield, CA**	-3.01	64
23420	**Fresno, CA**	-3.51	65

Table A2: Mortgage Origination Interacted with Supply Elasticity

We report estimates of the relationship of RESIDUAL BOOK VALUE on annual % change in mortgage origination. Panel A shows the results with all MSAs. Columns (1)-(2) report the results on annual % Change in Total Mortgage Origination. Column (1) shows result for the period of 2000 to 2006 and column (2) shows the result for the period of 2007 to 2010. We control for RESIDUAL BOOK VALUE, Supply Elasticity, the quintile dummies of Supply Elasticity, High Income Growth Dummy, High POP Growth Dummy and include state fixed effects. RESIDUAL BOOK VALUE is the residual of log total book value on the quadratic form of log total income using years from 1996 to 2000. High Income Growth Dummy is an indicator for MSAs with above median income growth rate during 1990 to 2000. High POP Growth Dummy is an indicator for MSAs with above median population growth during 1990 to 2000. For brevity, coefficients on quintile dummies of Supply Elasticity are not reported. Columns (3)-(4) report the results on annual % Change in Second Home Mortgage Origination. Column (3) shows result for the period of 2000 to 2006 and column (4) shows the result for the period of 2007 to 2010. Panel B shows the results only with large MSAs whose average population is above 750000. We use MSA population in 2000 as a weight for all regression results. The table reports point estimates with t-statistics in parentheses. We use robust standard errors. ***, **, * denotes 1%, 5%, and 10% statistical significance.

Panel A: All MSAs	% Change in Total MTG Origination		% Change in Second Home MTG Origination	
	2000 - 2006	2007 - 2010	2000 - 2006	2007 - 2010
Variables	(1)	(2)	(3)	(4)
RESIDUAL BOOK VALUE	-0.0293***	0.0357**	-0.0374*	0.0242*
	(-2.888)	(2.572)	(-1.774)	(1.741)
Supply Elasticity	-0.00956	0.000142	-0.0173	0.0475
	(-1.360)	(0.0112)	(-0.584)	(1.036)
High Income Growth Dummy	-0.0428***	0.0457**	-0.0346	-0.00308
	(-3.992)	(2.310)	(-1.413)	(-0.104)
High POP Growth Dummy	0.0559***	-0.0592***	-0.00261	-0.0350
	(3.401)	(-2.620)	(-0.0498)	(-1.321)
RESIDUAL BOOK VALUE × Elasticity Dummy (2/5)	0.0240	-0.00952	0.0708	-0.00325
	(1.470)	(-0.625)	(1.058)	(-0.237)
RESIDUAL BOOK VALUE × Elasticity Dummy (3/5)	0.0141	-0.0112	-0.0310	-0.00624
	(1.197)	(-0.728)	(-0.486)	(-0.417)
RESIDUAL BOOK VALUE × Elasticity Dummy (4/5)	0.0306***	-0.0366**	0.0293	-0.0345*
	(2.841)	(-1.990)	(1.409)	(-1.816)
RESIDUAL BOOK VALUE × Elasticity Dummy (5/5)	0.0173	-0.0326**	-0.00682	-0.0352
	(1.360)	(-2.216)	(-0.144)	(-1.165)
Observations	277	277	277	277
R-squared	0.780	0.765	0.336	0.484
State FE	Yes	Yes	Yes	Yes

Panel B: Large MSAs	% Change in Total MTG Origination		% Change in Second Home MTG Origination	
	2000 - 2006	2007 - 2010	2000 - 2006	2007 - 2010
Variables	(1)	(2)	(3)	(4)
RESIDUAL BOOK VALUE	-0.0589**	0.0595	-0.0429	0.0329
	(-2.264)	(1.536)	(-1.299)	(1.056)
Supply Elasticity	0.0315	-0.00839	0.0248	0.0593
	(0.484)	(-0.0984)	(0.245)	(1.047)
High Income Growth Dummy	-0.0523*	0.0751***	-0.0848**	0.0210
	(-2.078)	(3.214)	(-2.550)	(0.771)
High POP Growth Dummy	0.0502	-0.0259	0.00498	0.0495
	(1.200)	(-0.441)	(0.0936)	(0.970)
RESIDUAL BOOK VALUE × Elasticity Dummy (2/5)	0.0257	0.00455	0.0374	0.0149
	(0.691)	(0.0800)	(0.777)	(0.377)
RESIDUAL BOOK VALUE × Elasticity Dummy (3/5)	-0.00570	0.0218	-0.0386	0.0448
	(-0.145)	(0.315)	(-0.771)	(0.954)
RESIDUAL BOOK VALUE × Elasticity Dummy (4/5)	0.0203	0.000406	0.00700	0.0237
	(0.685)	(0.00925)	(0.180)	(0.668)
RESIDUAL BOOK VALUE × Elasticity Dummy (5/5)	0.0826	-0.101	0.0969	-0.126
	(1.269)	(-1.258)	(0.988)	(-1.718)
Observations	65	65	65	65
R-squared	0.980	0.963	0.981	0.972
State FE	Yes	Yes	Yes	Yes

Internet Appendix of: When Real Estate is the Only Game in Town

Hyun-Soo Choi* Harrison Hong† Jeffrey D. Kubik‡
Jeffrey P. Thompson§

This Draft: Dec 24, 2015

Abstract

This Internet appendix shows the robustness results of our household portfolio analysis.

*Singapore Management University and SKBI (e-mail: hschoi@smu.edu.sg)
†Princeton University, NBER and CAFR (e-mail: hhong@princeton.edu)
‡Syracuse University (e-mail: jdkubik@maxwell.syr.edu)
§Federal Reserve Board (e-mail: jeffrey.p.thompson@frb.gov)

1. Robustness of FICO Difference-in-Differences Results

We have tried several alternative specifications to show that the FICO difference-in-differences results are robust. In our basic difference-in-differences specification shown in Table 13, we compare households living in a MSA who live in the top and bottom quartile of the zip code FICO distribution. We omit households who live in the middle half of the distribution. We show in Table A3 the results when we include those households. The specification is identical to the specification shown in column (2) of Table 13, except that all households are included. Instead of only interacting log RATIO with a dummy for a household living in a top quartile FICO zip code, we interact log RATIO with three indicators for the household living in one of the top three quartiles of the zip code FICO distribution. The estimates suggest that there is no effect of log RATIO for households living in the second lowest quartile compared to the lowest. The negative effect of log RATIO becomes large in absolute value only for households in the third and especially the top quartile of the zip code FICO distribution. However, the coefficients are imprecise because we have added a number of additional quartiles to our baseline specification.

Up to this point, we have split households based on the FICO scores of their neighborhoods. But we could interact log RATIO with other household characteristics such as racial status or educational attainment. We next examine how our difference-in-differences results are affected by including interactions of log RATIO with whether the household is white. The results are presented in Table A4. In column (1), the coefficient on this interaction is negative and statistically different from zero, suggesting that the only-game-in-town effect is larger for whites than non-whites. In column (2), we only include the interaction of log RATIO with an indicator for the household having at least some sort of college degree. The coefficient on the interaction is negative, indicating that the only-game-in-town effect is larger for more educated households. But the coefficient on the interaction is small and imprecise.

In column (3), we run a horserace by including all three interactions with log RATIO (High FICO, White Household, and High Education). The interaction of High FICO Indicator and log RATIO is robust to the inclusion of these other interactions; the coefficient is very similar to previous estimates. However, the coefficients on the other two interactions are small and imprecise. These results point to why we emphasize low versus high FICO status in our empirical specifications.

In our framework, we do not expect to see much of a relationship between our only-game-in-town variable driving purchases of vacation homes because such homes are typically consumption decisions and not investment decisions. In Table A5, we present the results showing the relationship between log RATIO and vacation home, both the logit specification in columns (1)-(2) for whether or not a household owns a vacation homes and the tobit specification with holdings as the dependent variable in columns (3)-(4). Of course, wealth and income matter. But the coefficient of interest on log RATIO is either now insignificant or of the wrong sign. Also the High FICO indicator is less significant than before. Similar conclusions hold for when we consider primary homes as opposed to vacation homes as the dependent variable of interest. These results are omitted for brevity and are available from the authors.

Finally, recall that we found that even households in high FICO zip codes tilt toward investment homes compared to households in low FICO zip codes in low RATIO MSAs. We now show that they also tilt away from stocks. We show the results of the difference-in-differences specification from Table 13 but now using as the dependent variable stock holdings rather than whether a household has investment real estate. These tobit results are shown in Table A6.[1] We concentrate on comparing households in very high and low FICO zip codes: the top and bottom 10% of the zip code distribution. The coefficient of interest

[1] Tobits including MSA $\times$ Year effects do not converge. Therefore, we report the estimates with MSA and Year effects separately as opposed MSA interacted with Year. OLS estimates where we include MSA $\times$ Year effects are almost identical to the results shown in Table A6.

is in front of log RATIO × High FICO Zip Code. The coefficient is 6.175 with a t-statistic of 2.11. Households in high FICO zip codes tilt more toward stocks compared to households how FICO zip codes in high RATIO MSAs than low RATIO MSAs. Another way of saying this is that these households in low RATIO MSAs do not tilt toward stocks but as we showed above tilt toward investment homes. The magnitude of the interaction term suggests that all of the sensitivity of household behavior in stock holding to log RATIO is concentrated in households living in very high FICO zip codes.

1.1. Local Bias of Portfolios in SCF Sample

Although we have cited substantial evidence from the literature that there is local bias in purchases of investment homes, we also want to examine whether we can find evidence of this bias in the SCF. The SCF does not provide information on the location of a household's real estate investments. But we can infer the location of the households' real estate holdings using the 2007 SCF. Households in that wave of the SCF were re-interviewed two years later. Therefore, we can create a short panel that measures the change in a household's assets between 2007 and 2009.

For each MSA with a population greater than 750,000, we sum the value of investment real estate, vacation homes and stock assets of all the households in that MSA in 2007 and 2009. We then calculate the percentage change of those assets over those years. We can correlate those percentage changes in assets with the local return of those assets over the same time period. This will give us a measure of how local these investment are.[2]

The results of these regressions are reported in Table A7. In column (1), we estimate the localness of households' investment real estate assets. We regress at the MSA level the percentage change in investment real estate assets of a MSA on the return to housing in that MSA over the same time period. This MSA housing return is calculated based on FHFA housing price indices. The coefficient on the MSA housing return is about one and

[2] Note that we only have one cross-section and hence there is no need to control for price changes in the overall market.

statistically different from zero. Changes in local housing prices strongly predict changes in investment real estate assets, suggesting that a large percentage of these investment real estate investments must be local.

In column (2), we estimate a similar regression specification, except that the dependent variable is a measure of the percentage change in vacation home assets in a MSA instead of investment real estate. The relationship between this measure of vacation home holdings and local housing returns is much smaller. The coefficient on MSA housing returns is about 0.27 and statistically insignificant. Like previous work, we find that vacation home investments are substantially less local than investment real estate.

Finally, we use a similar methodology to measure the localness of stock investments of households in the 2007 SCF. In column (3), we regress the percentage change of stock holdings within a MSA on a measure of the stock returns of local stocks over the same time period. The coefficient on the MSA local stock return is 0.47 with a t-statistic of 1.53. Therefore, our results suggest that the localness of stock investment is not as large as investment real estate but somewhat larger than vacation homes.

1.2. Alternative Explanations: Accounting for Local Housing Affordability and Differential Economic Growth Across MSAs

We now consider two alternative explanations for our findings. The first is that the log RATIO of a MSA might be positively correlated with housing affordability in that MSA. We might simply be seeing less investment homeownership for households in high FICO zip codes in MSAs with lots of firms just because homes are more expensive there compared to stocks. This would bias our estimates of the relationship between log RATIO and investment home ownership upward. We show that accounting for home prices by interacting housing affordability at the MSA level with log RATIO or controlling for home prices at the zip code level actually makes our results stronger.

Therefore, it is not clear whether the relationship we find between the relative holdings of real estate within a MSA are caused by log RATIO differences across MSAs or real estate price differences. Our best guess is that this concern would bias our diff-in-diff results toward zero. In a high log RATIO MSA that also has high real estate prices, it is probably harder for households living in low FICO zip codes to afford real estate than households living in a high FICO zip code. This would mean that households in high FICO zip codes in high log RATIO/real estate price MSA would own more investment real estate compared to low FICO zip code households living in that MSA. But we find that the opposite is true in Table 13.

Because one can tell other stories about the potential bias of real estate prices, we estimate our difference-in-difference model controlling for the affordability of real estate in MSAs to see how those controls change our estimates. The results of this analysis are presented in Table A8. We deal with this alternative explanation in two ways. The first way is to use the specification in column (2) of Table 13, except that we also add to the OLS specification a measure of the affordability of housing at the MSA level interacted with the High FICO Zip Code indicator. We are allowing high FICO households to behave differently compared to low FICO households across MSAs with different housing affordability. The additional control variable is not statistically significant. The coefficient on the interaction term of interest, log RATIO interacted with a dummy for living in a high FICO zip code, is very similar to before when these affordability controls are not included. This suggests that our initial difference-in-differences results are not being driven by housing affordability.

In columns (2) and (3), we control of home prices at the zip code level. Because we do not have home prices at the zip code level for our entire sample, we first show in column (2) that using this smaller sample does not importantly change our standard difference-in-differences estimate. In column (3) we add the zip code price control. The magnitude of the coefficient on our interaction term of interest increases somewhat in absolute value when we add this control, providing additional evidence that a correlation between log RATIO and real estate prices locally if anything might bias our results toward zero.

The second alternative explanation is that MSAs with lots of local firms might simply have higher economic growth rates going forward and stocks are much more sensitive to this economic growth. If households in high FICO zip codes are better at understanding this than households in low FICO zip codes, then it is only in MSAs with many firms that we would see households in high FICO zip codes tilt toward stocks.

To address this concern, we can add to our difference-in-differences specification interaction terms for MSA Unemployment and High FICO Zip Code and also Past per capita MSA Income growth interacted with High FICO Zip Code. This is shown in column (1) of Table A9. The idea again is that these interaction controls allow high FICO households to behave differently relative to low FICO counterparts across MSAs with different unemployment and per capita income growth. If this alternative explanation is correct, then this ought to be driving our results and not log RATO $\times$ High FICO Zip Code. Again, we find no change in our coefficient of interest.

One might object that current unemployment or past income growth are not good indicators of economic growth. That is, investors are expecting future growth. We can redo our estimates but use future unemployment and future per capita income growth instead. This is shown in column (2) of Table A9; there is no difference in the results.

We view these two alternative explanations: housing is not affordable in high FICO neighborhoods and high RATIO MSAs have more economic growth and high FICO investors recognize this, as being the most plausible confounders of our difference-in-differences estimates. But after controlling directly for these alternatives, we still obtain very similar results.

A third alternative might be segmented credit markets. But if anything, one might think that living in high FICO neighborhoods might make it easier for households for households there to borrow and buy more, rather than less, investment homes. In any event, during sample period, 2002-2010, banking markets are by and large integrated (Landier et al. (2013)).

References

Landier, A., D. Sraer, D. Thesmar, et al. 2013. Banking deregulation and the rise in house price comovement. Tech. rep., Institut d'Économie Industrielle (IDEI), Toulouse.

Table A3: The Difference in Investment Home Ownership Between High and Low Credit Risk Households Within a MSA by log RATIO: Using the Entire Sample and FICO Quartiles

We report estimates of the OLS relationship of log RATIO on a family's holdings of investment real estate, splitting up the effect of log RATIO for households living in various credit risk zip codes. The sample consists of all households. The dependent variable is an indicator for whether the family owns investment real estate. The independent variables are the same as in Table 6, except that instead of only including a dummy variable for a household living in the highest quartile FICO zip code, we include dummies for all quartiles (omitting the lowest). The table reports OLS point estimates with t-statistics in parentheses. All the standard errors are clustered at the MSA level. ***, **, * denotes 1%, 5%, and 10% statistical significance.

Variables	(1) HAVE INVESTMENT REAL ESTATE
log RATIO × Highest Quartile FICO Zip Code	-0.0172
	(-1.61)
log RATIO × Third Quartile FICO Zip Code	-0.0051
	(-0.93)
log RATIO × Second Quartile FICO Zip Code	-0.0007
	(-0.12)
Highest Quartile FICO Zip Code	-0.0241
	(-1.61)
Third Quartile FICO Zip Code	-0.0107
	(-0.98)
Second Quartile FICO Zip Code	-0.0013
	(-0.14)
log HOUSEHOLD INCOME	0.0395***
	(9.54)
log HOUSEHOLD NET WORTH	0.015***
	(8.67)
UNATTACHED FEMALE	-0.0215**
	(-2.22)
Other Household Controls	Yes
Year Effects	—
Year × MSA Effects	Yes
Observations	89956

Table A4: The Difference in Investment Home Ownership Across Race, Education and Credit Risk Groups Within a MSA By log RATIO

We report estimates of the OLS relationship of log RATIO on a family's holdings of investment real estate, splitting up the effect of log RATIO by race and education. In the first two columns, the sample consists of all households. In the last column, the sample consists of families who live in a zip code in the highest and lowest 25% of the credit risk distribution (measured by average FICO scores). The dependent variable is an indicator for whether the family owns investment real estate. The independent variables are the same as in Table 6, except that in column (1) we add an interaction term of log RATIO and an indicator for the household being white. In column (2), we instead add an interaction of log RATIO and an indicator that the household has a bachelors degree or better. The final column adds both of the previous interactions and the log RATIO and high FICO zip code interaction. The table reports OLS point estimates with t-statistics in parentheses. All the standard errors are clustered at the MSA level. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	(1)	(2)	(3)
Variables	HAVE INVESTMENT REAL ESTATE		
log RATIO × White Household	-0.0096**		0.0059
	(-1.98)		(0.4)
log RATIO × Bachelor Degree or Better		-0.0023	-0.0045
		(-0.27)	(-0.24)
log RATIO × High FICO Zip Code			-0.0243*
			(-1.93)
High FICO Zip Code	-0.0055	-0.0056	-0.0337
	(-0.50)	(-0.51)	(-1.43)
log HOUSEHOLD INCOME	0.0395***	0.0394***	0.0441***
	(9.54)	(9.52)	(7.4)
log HOUSEHOLD NET WORTH	0.015***	0.015***	0.014***
	(8.62)	(8.57)	(6.17)
UNATTACHED FEMALE	-0.0215**	-0.0218**	-0.015
	(-2.23)	(-2.26)	(-0.95)
Other Household Controls	Yes	Yes	Yes
Year Effects	—	—	—
Year × MSA Effects	Yes	Yes	Yes
Observations	89956	89956	41845

Table A5: Vacation Home Ownership, % Share of Vacation Home and log RATIO

We report estimates of the relationship between log RATIO and measures whether a family owns of a vacation home. The dependent variable in columns (1) through (2) is an indicator that a family owns a vacation home. The coefficients in those columns are from a logit model. The average marginal effect is in brackets. The dependent variable in columns (3) through (4) is the share the familys assets invested in a vacation home. The coefficients in those columns are from a tobit model. The independent variables include log RATIO, the log ratio of the total book value of firms headquartered in a MSA to the income in that MSA as in Hong, Kubik, and Stein (2008) and log HOUSEHOLD INCOME, the usual income of the family and log HOUSEHOLD NET WORTH, the wealth of family (not including real estate). Due to the MSAs with RATIO equal to zero, we add 0.00001 to RATIO before we take log. High FICO Zip Code is a dummy for the household living in a zip code in the top quartile of the average FICO score distribution. UNATTACHED FEMALE is an indicator that the family is headed by an unattached woman. Other Household Controls include age, race, and education-level of the head of the family, family structure, and a linear and square term of the number of people in the family. The table reports point estimates with t-statistics in parentheses. All the standard errors are clustered at the MSA level. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	(1)	(2)	(3)	(4)
Variables	HAVE VACATION HOME		% VACATION HOME	
log RATIO	0.0417	0.0635*	0.0061*	0.0071**
	(1.08)	(1.76)	(1.73)	(2.1)
	[.0021]	[.0032]		
High FICO Zip Code	-0.0738	-0.104	-0.0125	-0.0183*
	(-0.97)	(-1.31)	(-1.28)	(-1.81)
log HOUSEHOLD INCOME	0.273***	0.276***	0.0397***	0.0398***
	(5.38)	(5.48)	(5.97)	(5.93)
log HOUSEHOLD NET WORTH	0.434***	0.433***	0.0425***	0.042***
	(12.96)	(12.37)	(9.79)	(9.81)
UNATTACHED FEMALE	0.315	0.315	0.029	0.03
	(1.5)	(1.49)	(1.17)	(1.24)
Other Household Controls	Yes	Yes	Yes	Yes
Year Effects	Yes	—	Yes	—
Year × Census Division Effects	No	Yes	No	Yes
Observations	89956	89956	89956	89956

Table A6: The Difference in Stock Holdings Between High and Low Credit Risk Households Within a MSA by log RATIO

We report estimates of the relationship of log RATIO on a family's holdings of stock, splitting up the effect of log RATIO for households living in low and high credit risk zip codes. The sample consists of families who live in a zip code in the highest and lowest 10% of the credit risk distribution (measured by average FICO scores). The dependent variable is the log of a family's holdings in stocks. The independent variables are the same as in Table 4 except for a couple of differences. log RATIO × High FICO Zip Code is the interaction of that indicator and log RATIO. MSA and Year Effects are also included in the specification. The coefficients are from a tobit model. The table reports point estimates with t-statistics in parentheses. All the standard errors are clustered at the MSA level. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	(1)
Variables	log FAMILY HOLDINGS in STOCKS
log RATIO × High FICO Zip Code	6.175**
	(2.11)
log RATIO	-0.159
	(-0.04)
High FICO Zip Code	1.381
	(0.53)
log HOUSEHOLD INCOME	2.329***
	(3.89)
log HOUSEHOLD NET WORTH	1.085**
	(2.39)
UNATTACHED FEMALE	0.223
	(0.09)
Other Household Controls	Yes
Year and MSA Effects	Yes
Observations	7275

Table A7: Measuring Local Bias in the 2007 SCF

We report OLS estimates of the relationship between the percentage change of households' holdings of different assets in a MSA over time and the return on housing in that MSA over the same time period. The sample consists of MSAs with a population greater than 750,000 with households in the 2007 SCF. The dependent variable in column (1) is the percentage change of investment home assets between 2007 and 2009 of 2007 SCF households within a MSA. The dependent variable in column (2) is a similar measure using the change in vacation home assets of households in the 2007 SCF. The dependent variable in column (3) is the same percentage change but for stocks. The independent variables are a constant and a measure of the return to housing in a MSA between 2007 and 2009 using FHFA house price information in the first two columns. In column (3), the independent variable of interest is the return to local stocks over that time period. All regressions are weighted by the amount of the asset held in the MSA in 2007. t-statistics are in parentheses. Robust standard errors are estimated. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	(1) Percentage Change Investment Homes	(2) Percentage Change Vacation Homes	(3) Percentage Change Stock Holdings
Constant	-0.08	-0.23	-0.37***
	(-0.52)	(-1.23)	(-4.20)
MSA Housing Return	1.14**	0.27	
	(2.19)	(0.29)	
MSA Local Stock Return			0.47
			(1.53)
Observations	40	33	41

Table A8: The Difference in Investment Home Ownership Between High and Low Credit Risk Households Within a MSA by log RATIO: Controlling for Housing Affordability

We report estimates of the relationship of log RATIO on a family's holdings of investment real estate, splitting up the effect of log RATIO for households living in low and high credit risk zip codes. The sample consists of families who live in a zip code in the highest and lowest 25% of the credit risk distribution (measured by average FICO scores). The dependent variable is an indicator for whether the family owns investment real estate. The independent variables are the same as the model shown in column (2) Table 6 except for a couple of additional controls for housing prices. In column (1), the specification includes the Housing Affordability Index value of the MSA of the household and the interaction of this index with High FICO Zip Code. The specifications in columns (2) and (3) only include observations that we have a measure of the median price of homes for that zip code (log HOUSE PRICE ZIP CODE). The table reports point estimates with t-statistics in parentheses. All the standard errors are clustered at the MSA level. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	Full Sample	Zip Code Home Price Sample	
	(1)	(2)	(3)
Variables	HAVE INVESTMENT REAL ESTATE		
log RATIO × High FICO Zip Code	-0.0232**	-0.0281**	-0.0305**
	(-2.07)	(-2.20)	(-2.29)
log RATIO	—	—	—
High FICO Zip Code	-0.0427*	-0.0348	-0.0622**
	(-1.81)	(-1.39)	(-2.06)
Housing Affordability Index	—		
High FICO Zip Code × Housing Affordability Index	-0.0002		
	(-0.21)		
log HOUSE PRICE ZIP CODE			0.034*
			(1.88)
log HOUSEHOLD INCOME	0.0448***	0.0449***	0.0429***
	(7.33)	(6.9)	(7.08)
log HOUSEHOLD NET WORTH	0.0133***	0.0178***	0.0178***
	(6.36)	(6.5)	(6.74)
UNATTACHED FEMALE	-0.0137	-0.0193	-0.0201
	(-0.86)	(-0.96)	(-1.01)
Other Household Controls	Yes	Yes	Yes
Year × MSA Effects	Yes	Yes	Yes
Observations	41845	27586	27586

Table A9: The Difference in Investment Home Ownership Between High and Low Credit Risk Households Within a MSA by log RATIO: Controlling for MSA Economic Conditions

We report estimates of the relationship of log RATIO on a familys holdings of investment real estate, splitting up the effect of log RATIO for households living in low and high credit risk zip codes. The sample consists of families who live in a zip code in the highest and lowest 25% of the credit risk distribution (measured by average FICO scores). The dependent variable is an indicator for whether the family owns investment real estate. The independent variables are the same as the model shown in column (2) Table 6 except for a couple of additional controls for MSA economic conditions. In column (1), the specification includes the unemployment value of the MSA of the household and the interaction of this index with High FICO Zip Code, and the past three year growth rate of per capita income of that MSA and the interaction of this value with High FICO Zip Code. In column (2), the specification includes forward looking values of these two variables: the unemployment rate of the MSA three years later and the growth rate of per capita income three years in the future. The table reports point estimates with t-statistics in parentheses. All the standard errors are clustered at the MSA level. ***, **, * denotes 1%, 5%, and 10% statistical significance.

	(1)	(2)
Variables	HAVE INVESTMENT REAL ESTATE	
log RATIO x High FICO Zip Code	-0.0375**	-0.0358**
	(-2.45)	(-2.21)
log RATIO	—	—
High FICO Zip Code	0.0464	-0.0347
	(0.44)	(-0.82)
MSA Unemployment	—	
MSA Unemployment x High FICO Zip Code	-0.01	
	(-0.93)	
Past Per Capita MSA Income Growth	—	
Past Per Capita MSA Income Growth x High FICO Zip Code	-0.375	
	(-1.14)	
Future MSA Unemployment		—
Future MSA Unemployment x High FICO Zip Code		-0.0009
		(-0.14)
Future Per Capita MSA Income Growth		—
Future Per Capita MSA Income Growth x High FICO Zip Code		-0.106
		(-0.56)
log HOUSEHOLD INCOME	0.0426***	0.0426***
	(6.93)	(6.93)
log HOUSEHOLD NET WORTH	0.0133***	0.0132***
	(5.86)	(5.74)
UNATTACHED FEMALE	-0.0215	-0.0212
	(-1.39)	(-1.37)
Other Household Controls	Yes	Yes
Year x MSA Effects	Yes	Yes
Observations	41845	41845