Investors in Housing Markets: Comparing Two Booms[*]

Carlos Garriga[†], Pedro Gete[‡], Athena Tsouderou[§]

December 2023

Abstract

We conduct a large-scale data analysis to compare housing investors in the periods before and after the Global Financial Crisis (GFC). Post-GFC investors differ from pre-GFC investors in several key ways: They are less likely to use leverage, are residents in wealthier and more educated areas, and are more sophisticated. They have a buy-and-hold strategy, are less likely to sell in response to capital gains and more likely to hold properties longer in response to rental yields. These differences have implications for market liquidity and monetary policy transmission sensitivity. Moreover, post-GFC investors are less susceptible to behavioral biases.

Keywords: Residential investment, rental yields, capital gains, mortgages, biases, housing, flipping.

1

1 Introduction

The housing boom of the 2000s, and its subsequent bust, provide a severe example of the significant role of housing investors in the United States economy. According to the literature, most housing investors during the 2000s boom, so-called *flippers* or *speculators*, were susceptible to behavioral biases and had a strategy of maximizing short-term capital gains from frequent buying and selling. Credit markets provided sufficient funding to investors to engage in high volume of transactions usually financed by rolling over high levels of debt. When the housing market collapsed in 2007, housing investors were among the hardest hit, with many of them being over-leveraged and having to sell their properties at a loss. This contributed to the severity of the housing crisis and the Global Financial Crisis (GFC). In this paper we show stark differences between the characteristics and strategies of housing investors pre- and post-GFC. These changes in the investors have reduced the possibility of a new housing market collapse. Also, they have important implications for market liquidity, the transmission of monetary policy and systemic risk mitigation.

We use detailed housing transaction data and track the purchase and sale of each property by investors. In addition, we utilize information from the name and the primary residence of investors and the location and characteristics of the houses they buy. We uncover the following findings: 1) Post-GFC investors have longer holding periods than pre-crisis investors; 2) They are less sensitive to the capital gains they would realize when deciding to sell; 3) They have higher sensitivity to rental yields, in line with their preference for a long horizon of investment; 4) They are wealthier, more educated and more sophisticated than their pre-crisis counterpart; 5) Post-GFC investors are significantly less likely to finance their purchases with a mortgage.

Our analysis is based on the universe of housing transaction deeds between 2000 and 2017 across the whole U.S. We focus on single-family homes as they are the most common residential unit transacted. In total we have 60 million transactions by homeowners and investors. We use the buyer name and the number of properties they purchase to identify investors. We keep track of when each property was bought and sold again to compute holding periods, using the unique identifier of each property. We use the registered lien against the purchased properties in the deeds records to identify the use of credit. We match the transactions with property characteristics from tax assessor data. Furthermore, using census tract, county and Metropolitan Statistical Area (MSA) geographical units, we merge the deeds data with house prices and rents, economic trends and demographic characteristics from different data sources.

We focus on *retail investors*, defined as those individuals who bought two or more houses in a two-year period within the same MSA using their personal names or a legal entity (for

example an LLC, LP, trust or corporation). It is unlikely to have more than two homes in the same MSA for leisure purposes, particularly when considering a vacation home. The definition of retail investor excludes the 26 largest "Wall Street landlords" and their subsidiaries.[1] This exclusion allows us to focus on comparing small and medium investors pre- and post-crisis, as they account for nearly all investors' transactions in single-family homes.

We study two periods with similar size and duration of house price appreciation. Between 2001 and 2006 (pre-GFC), U.S. real housing prices increased by around 45%, and between 2012 and 2017 (post-GFC) by around 40% as Figure 1 shows. The similarity between these two booms allows us to compare the characteristics of housing investors in each of them.

We find that post-GFC investors live in census tracts with higher real median incomes and higher housing wealth, standardised by the real median income and housing wealth of the MSA containing the census tract. Moreover, they live in census tracts with a higher percentage of the population holding at least a bachelor's degree or at least a master's degree than the pre-GFC investors. Thus, post-crisis housing investors are wealthier and more educated than pre-crisis investors. Additionally, as an indication of investment savvy and sophistication, the majority of new investors create a trust, limited partnership, or limited liability company to purchase their properties.

Almost all post-GFC investors in the single-family housing market are new investors, as they have different names and mail addresses from the pre-GFC investors. This is consistent with a new type of investor entering the market and becoming the dominant player, rather than the same pre-GFC investors learning from the crisis or changing their preferences as they react to new market conditions.

We estimate logistic regressions to infer the investors' preferences towards holding or selling the properties that they own. Post-crisis investors are less sensitive to sell in the short-term based on capital gains than pre-crisis investors. Specifically, for one standard deviation higher capital gains from house price appreciation, the probability of selling within two years is 4.0% higher for pre-crisis investors and 0.3% higher for post-crisis investors. Post-crisis investors have substantially longer holding periods than pre-crisis investors, making them long-term investors. The preference for a long horizon of investment is manifested in large sensitivity to rental yields. For one standard deviation higher rental yield, the probability of selling within two

---

[1]Pre-GFC the housing investors in our sample are mostly individuals, having their personal name in the deed. Post-GFC there is an emergence of large institutional capital investors referred to as "Wall Street landlords" (see, for example, Gurun et al. 2023). Despite the media exposure in the popular press, they only account for less than 2% of the transactions in any given year, and their purchases are very concentrated in a number of selected locations. To make the exclusion accurate, we cross check their primary names with all the subsidiary names using SEC filings. The Appendix provides additional details about the data.

years decreases by 1.8% for pre-crisis investors and by 8.2% for post-crisis investors.

Our data allows to study the use of mortgage credit, as deeds record any registered lien against the purchased property. In the 2000s boom, 65% of house purchases by retail investors were financed with a mortgage, but this percentage drops to 23% for post-crisis investors. The pre- and post-GFC average loan-to-value ratio of retail investors was 0.83 and 0.81, respectively. The analysis using t-test indicates that the loan-to-value difference is statistically significant, albeit small. The extensive margin of credit is what clearly differentiates the post-GFC investors from the pre-GFC, as they are more likely to use cash than credit to complete housing transactions.[2]

The previous findings have important implications on many dimensions. There is a large literature showing that wealthier, more educated, and generally more sophisticated investors often make better financial decisions and exhibit fewer investment biases.[3] Thus, our findings that the post-crisis investors are more educated, wealthier, and more sophisticated suggest that the new investors are less prone to the behavioral biases studied in the literature for the early 2000s boom. These new investors are less likely to form irrational bubbly expectations by extrapolating from past price increases. Moreover, since they rely less on debt, they will default less as housing prices fall. Especially, they will mitigate the amplification caused by default with non-recourse mortgages. For instance, as housing prices collapsed during the GFC, under-water investors defaulted on their loans. In addition, the transmission channel of monetary policy will operate differently for these new housing investors. The new investors are less sensitive to increases in mortgage rates as they have less debt. Furthermore, the liquidity properties of housing markets will also change as the new investors are long-term, more sensitive to rental yields and less reactive to capital gains. There will be less inventory for sale during boom periods, as buy-and-hold investors decrease transaction volumes.

Monetary policy is the most likely explanation for the arrival of these investors. They were attracted to housing markets while searching for yield when returns from deposits were very low. Daniel, Garlappi, and Xiao (2021) find that a low-interest-rate monetary policy increases investors' demand for high-dividend stocks. As rents are mostly stable, housing becomes an investment asset and a close substitute to safe yield-earning investments (see for example, Demers and Eisfeldt 2022, Eichholtz et al. 2021). Korevaar (2023) provides evidence that investors reached for yield using data from 18th-century Amsterdam. One could expect that as

---

[2]See Gete and Reher (2016) for a model characterizing the extensive and the intensive margins of mortgage credit.

[3]See for example, Agnew (2006), Dhar and Zhu (2006), Kumar (2009), Calvet, Campbell, and Sodini (2009), Bose, MacDonald, and Tsoukas(2015), Cronqvist and Siegel (2015) and Gomes, Haliassos, and Ramadorai (2021) among others.

monetary policy normalizes and housing offers less opportunities for search for yield, post-crisis investors could resemble more the pre-crisis investors. However, the current environment with high and persistent inflation provides an additional motive to continue using housing holdings as a hedge against it. Housing as an asset class has performed extremely well during the pandemic, especially now that the role of bonds to insure stock volatility has changed with strong positive correlation between the two.[4]

Our analysis contributes to the active literature studying investors in housing markets.[5] Our findings show that post-crisis investors have clear differences with the ones documented by the literature analyzing the 2000s housing boom. For example, DeFusco, Nathanson, and Zwick (2020) document that much of the variation in housing volume during the 2000s boom-bust came from the rise and fall in speculation. Bayer, Mangum, and Roberts (2021) show that speculation during the boom period was driven by short-horizon investors as a response to past housing price increases. Bhutta (2015) provides evidence that mortgage debt for real estate investors grew markedly during the 2000s housing boom. Garcia (2022) finds that counties with higher share of investor activity experienced higher house flipping rates pre-crisis. Haughwout et al. (2011) document a large growth in short-term speculators in the early 2000s. Piazzesi and Schneider (2009) document trading based on the increased belief in rising housing prices during the housing boom of the 2000s. The access to lax credit was important to sustain their investment strategy as documented by Albanesi, DeGiorgi, and Nosal (2012). There are clear investment biases for investors in the pre-GFC period. Chinco and Mayer (2016) provide evidence that out-of-town second-house buyers behaved like misinformed speculators. Another bias comes from their ability to forecast future housing price appreciation. Gao, Sockin, and Xiong (2020) provide evidence linking housing speculation during the pre-crisis housing boom to extrapolation behavior by speculators of past housing price changes. Glaeser and Nathanson (2017) show evidence that investors underreact to news because of behavioral biases or loss aversion. Loewenstein and Willen (2023) show that sources of growth in house prices varied from the 2000s boom that was plausibly driven by exuberant expectations, to the boom periods of 2010s and 2020s.

Our paper is part of a growing literature studying different aspects of post-crisis investors. The main focus of this literature has been to measure the impact of investors on house prices and rents. For example, Allen et al. (2017), Lambie-Hanson, Li, and Slonkosky (2022), Mills,

---

[4]In the last two decades, stocks and bonds exhibited almost no correlation or had a slight negative correlation around -8%. Since the beginning of 2022, there has been a dramatic increase in correlation between these two assets where the rolling 24-month correlation reached a 28-year high of 75%.

[5]A more established literature has studied foreign and out-of-town investment in the housing markets. See for example recent papers by Cvijanovic and Spaenjers (2021), Davids and Georg (2020), and Favilukis and Van Nieuwerburgh (2021).

Molloy, and Zarutskie (2019), and Garriga, Gete, and Tsouderou (2023) show that investors have increased house price growth, rent growth, and price-to-income ratios. Gurun et al. (2023) show that increased concentration of institutional landlords in some neighborhoods increases rents while reducing crime rates. Garriga, Gete, and Tsouderou (2023) document a connection between the change in investment strategy and monetary policy, via quantitative easing (QE), and the low return on safe assets. Using German data, Boddin et al. (2023) also document a housing portfolio channel related to QE, exploiting household-level and regional variation in the data. Relative to the literature, the analysis in this paper is the first to document the characteristics of these investors, in terms of wealth, education, leverage, or sensitivity to capital gains or rental yields using a unified framework and granular data.

The rest of the paper is organized as follows: Section 2 describes the data and the methodology to classify investors. Section 3 shows the investors' characteristics. Section 4 studies investors' sensitivity to capital gains and to rental cash-flow yields. Section 5 discusses the implications of the findings and Section 6 concludes. The online appendix has extra information about the variables and data sources.

2 Data and Definitions

The transaction data come from Corelogic. The Deeds Dataset contains all ownership transfers of single-family homes in the U.S., as recorded by the counties' deeds. We focus on the deeds from January 1st, 2000 to December 31st, 2017. We use extensive information from each deed, such as the name of the buyer, the name of the seller, the date of the transaction, the address of the property, the mailing address of the buyer, the price at which the property was sold and the loan amount on the property.

We merge the transaction data with property characteristics that come from tax assessors. The characteristics include the year the housing unit was built, total number of rooms, number of bedrooms, size of the house measured in square feet, and the year of the last renovation or update of the structure.

We use a rigorous methodology to classify retail investors from the transaction data. First, we filter out buyer names that are relocation companies, non-profit organizations, construction companies and national and regional authorities. Additionally, we filter out financial institutions such as banks, Ginnie Mae, the Government Sponsored Enterprises, mortgage loan companies and credit unions, and the state taking ownership of foreclosed properties. We define retail investors based on any of the following criteria: 1) The buyer of a given house is an

individual who has bought two or more houses in that given year or the previous year, within the MSA of the given house. Thus, individuals who purchase their main residence and a vacation home in a different MSA are not counted as retail investors. 2) The buyer is a legal entity (such as an LLC, LP, trust, or corporation) and is not one of the largest "Wall Street Landlords" who entered the market after the crisis.[6] We remove these large institutional investors to focus on retail investors when comparing the pre- and post-crisis periods. It is important to note that "Wall Street Landlords" are very small in the aggregate (Garriga, Gete, and Tsouderou 2023). We perform robustness checks and show that the presence of large institutional investors does not alter any of the findings presented in the next Sections.

One of the challenges is that the deeds dataset does not include demographic characteristics of buyers and sellers. To overcome this challenge we use the home address of the buyer as reported in the deed. This allows us to link each investor to the demographic and housing market data of the areas they live in. We obtain the data on the median income, education level, and owner-occupied house prices at the census tract from the Federal Financial Institutions Examination Council census and demographic data.

Moreover, we link the property address with data from various sources to obtain the evolution of housing prices, rents and economic trends in the areas where investors go. In particular, house prices at the zip code level and monthly frequency come from the Zillow Home Value Index for single-family homes. This index is designed to capture the value of a typical property across each zip code, not just the homes that sold, and it does so by utilizing information from the full distribution of homes in a given zip code. The housing rents also come from Zillow. The smallest geographical unit available for this series that goes back to the year 2000 is the MSA. We collect housing prices and rents with monthly frequency at the MSA level and calculate the rental yield as the ratio of 12 times the monthly rent over the price of the house.

Additional variables are population data from the U.S. Census Bureau, unemployment rate from the U.S. Bureau of Labor Statistics, and per capita income from the Bureau of Economic Analysis. More detailed description of the data sources is included in the online Appendix A.

The key variable to measure investors' presence is *the share of retail investors*, which we compute as the ratio of the number of single-family purchases by retail investors to the total number of single family purchases by investors and homeowners.

Tables 1 to 3 describe the main variables in the dataset for both periods that we study. We will discuss these tables in the next sections.

---

[6]We follow Amherst (2018) which estimated the number of institution-owned single-family rentals as of 2017 and classified the largest twenty-six investors.

3 Stylized Facts About Retail Housing Investors

This section discusses the new investors that arrived to the single-family housing market post-GFC. These investors differ from their pre-GFC counterparts in both their names and residential addresses. In relative terms, they have more education, income and wealth than the pre-GFC investors. Also, the new investors are more sophisticated and use less leverage.

3.1 Changes in Investors' Composition

We begin by investigating whether the housing investors post-crisis are the same investors than pre-crisis or if they are new. We focus on investors who bought properties in the years 2001-2004 and 2012-2015. These are the years of the two booms, including two additional years in each period 2001-2006 and 2012-2017, as we show in Figure 1. We focus on the purchases in those dates and track whether each property was sold in two years or not. Thus, we consider the full boom periods for the sale of the properties.

We show that the vast majority of the single-family investors post-GFC are new individuals or legal entities who were not in the market during the boom of the 2000s. To do so, we check whether the same investors are active in the market pre- and post-GFC. Specifically, we match the name of the buyer and MSA of the buyer address in the two periods. That is, we assume that if an investor is active in both periods, he or she will have the same name and will be living in the same MSA. Only 1.8% of the investors in the deeds have this commonality.

Furthermore, we investigate whether the names of the investors have changed. To do so, we match the address of the buyer in the two periods. We assume that if the same investor is active in the two periods, he or she will be using the same mailing address in the deeds. We find that 2.4% of the mailing addresses of the investors match pre- and post-crisis. We perform various robustness tests, for example using the longitude and latitude of the mailing address to do the matching, and still we find similar percentages of commonality. As we show later, the majority of post-crisis investors buy through a legal entity. This exercise shows that the new legal entities are registered in different addresses from the addresses of the investors pre-GFC. In Section 5 we discuss why this change happened and its important implications.

3.2 Investor Characteristics Post-GFC: More Educated, Wealthier and Sophisticated

In this subsection we compare key characteristics of retail investors pre- and post-crisis. We investigate whether the investors post-crisis have different social backgrounds in terms of education and wealth. This is important because low-income low-educated households are more prone to behavioral biases, for example, in what concerns forming expectations of housing price dynamics.

The initial inquiry is to identify the income level of retail investors across these two episodes. To do that, we compare the median income in the census tract of the listed residence address for the investor relative to the median income in the MSA in which the tract belongs. If the address has a higher median income than the median income of the MSA, we can indirectly infer that the investor has a relatively higher income. As such, ratios above one show that investors live in affluent neighborhoods of the MSA.[7] These values can be compared across locations by examining the distribution for each episode. Figure 2 documents the distribution of investors' income ratio for pre- and post-crisis investors. The distribution of the ratio has clearly shifted to the right for the post-crisis investors, highlighting that they tend to live in richer areas than pre-crisis investors.

We construct a proxy for wealth using the ratio of the median house price in the census tract of the investor's home to the median house price of the MSA where the investor lives. If the ratio shows that the median house value of the area where the investor lives is higher than the median of the MSA, one can infer that the investor's wealth is above the median. Figure 3 shows that the distribution of the ratio of house prices has shifted to the right, indicating that post-crisis investors tend to live in areas where homes are more expensive.

Both Figures 2 and 3 suggest that the new investors are wealthier or live in wealthier census tracts than the pre-crisis investors. This may indicate that they are not investing to speculate and accumulate wealth to become rich, but rather to place their savings in search of yield. We explore this conjecture can with additional detail by examining the use of leverage for transactions in both periods.

We note that, in general, there is high correlation between income and wealth, and both variables are also correlated with the level of education. Figure 4 shows the distribution of the share of the census tract population over 25 with a Bachelor's degree or higher (top panel) and

---

[7]The ratio of median family income in the census tract to the median family income in the MSA is used to define low-income community census tracts and opportunity zones. Our measure is inspired by this ratio.

the distribution of the share of the census tract population over 25 with a Master's degree or higher (bottom panel) in the census tracts of the investors' address. Investors in the post-crisis period live in areas with the distribution of these share shifted to the right.

Figures 2 to 4 provide a useful visual representation of the distributional differences between pre- and post-crisis investors. To quantify these differences, Table 1 summarizes the results of statistical tests of mean differences. The differences are statistically significant. Post-crisis investors reside in higher-income and wealthier census tracts with more educated residents.

The left panel of Figure 5 plots the percentage of total number of purchases of single-family homes by retail investors (flow of investment) during the years 2000-07. The solid line plots the percentage of legal entities and the dashed line the percentage of individual investors. The right panel plots the flow of investment from these two types of investors during the years 2010-17. Legal entities are small or medium-sized buyers who purchase homes through an LLC, LP, trust etc. Individual investors are individuals who purchase two or more houses within two years in the same MSA. As Figure 5 shows, during the boom of the 2000s the majority of purchases by retail investors (60.5%) came from individuals. However, during the 2010s boom the majority of purchases by retail investors (58.8%) came from legal entities.

Forming a legal entity requires a certain level of sophistication beyond an individual purchase, and signals the desire to use the legal form to be protected from unlimited liability. Thus, we find that post-crisis investors are more sophisticated. Table 1 confirms this finding as it shows that the flow of purchases by legal entities had a level increase of 19.2% post-crisis, which was statistically significant.

Overall, in this subsection we use the same data sources to compare two housing booms in the 2000s and 2010s, and we find that the investors in the single-family housing market have different socio-economic backgrounds across periods.

3.3   Cash vs. Credit Purchases

High leverage and real estate collateralization were key drivers of the Global Financial Crisis. As house prices collapsed and borrowers defaulted, this created a reinforcing negative loop of lower asset values and more defaults (Sufi and Taylor 2022). Several studies have documented that housing investors were highly leveraged in the years leading up to the crisis.

The data allows to identify the use of credit, as deeds record any registered lien against the purchased property. Table 2 shows that post-crisis investors used credit to finance their housing purchases less frequently (extensive margin) and slightly lower leverage ratios (intensive

margin) compared to pre-crisis investors. Specifically, 65% of house purchases by retail investors were financed with a mortgage in the 2000s, but only 23% for post-crisis investors. These are striking and sizable differences. Conditional on receiving credit to purchase a residential unit, the average loan-to-value ratio is only 2.1 percentage points lower, albeit statistically significant. This striking finding suggests that post-crisis investors were more likely to be savers looking for new investment opportunities than speculators, and provides additional clarification on the results summarized in Section 3.2. In other words, post-crisis investors were less likely to borrow to finance their housing purchases, but those who did borrow did so at slightly lower levels of leverage as pre-crisis investors. This suggests that post-crisis investors were more sophisticated and less likely to engage in the risky leveraged speculation that contributed to the financial crisis.

4   New Investment Strategies

The previous analysis uncovered that post-crisis investors had different socio-economic characteristics and were less likely to use credit when purchasing. In this section we study a consequence of such characteristics: their sensitivity to capital gains and rental yields. In other words, we ask whether the investors post-crisis are more likely, compared to the pre-crisis investors, to sell a house in the short-term, based on their potential capital gains and rental yield.

4.1   Investment Horizons

The literature has shown that pre-crisis investors had a very short-term horizon, one to three years, and they were looking to extract capital gains quickly from the booming market. We follow all the housing purchases by investors in the U.S. from the beginning of 2001 until the end of 2004, and from the beginning of 2012 until the end of 2015. We track which of the houses purchased were sold within two years of their purchase. Thus, we keep track of the market until the end of 2006 and 2017. We have therefore very good coverage of both housing booms.[8]

Figure 6 shows the distribution of house price growth in the counties where investors buy properties. The histogram compares these distributions for properties purchased in 2001-2004 versus properties purchased in 2012-2015. We can see that both distributions are similar, with

---

[8]We did lots of sensitivity tests. The results of this section are robust to the selection of pre- and post-crisis boom years.

the exception that there is more weight of investors post-crisis buying in areas with negative price growth.

Figure 7 shows the same distribution but for rental yield. As we noted before the rental yield is the annual rent over the price of the house. We see post-crisis investors buying mostly in locations with higher rental yields relative to pre-crisis investors. While pre-crisis the investments happened in areas with an average of 7.6% rental yield, post-crisis the investors targeted areas with an average of 9.4% rental yield. The new investors have placed significantly more weight in areas with rental yields around 9 to 12%. These histograms motivate the following analysis.

First, in Figures 8 and 9 we look at short-term sales by investors. Figure 8 plots the percentage of houses bought by retail investors in a particular year that were subsequently sold between 6 months and 3 years of their purchase.[9] The left panel plots this percentage for the houses bought during the pre-GFC boom, 2000-04, and the right panel during the post-GFC boom, 2010-15. We calculate the holding period as the number of days from the date the deed was signed to buy a particular house, until the date the following deed was signed to sell the house.

Figure 9 plots the holding duration of investors similar to Figure 8 but for those MSAs with the largest boom–bust cycles during the 2000s as documented by DeFusco, Nathanson, and Zwick (2020). Both figures show an astonishing drop in the percentage of investors investments that are short term, especially in those locations that according to DeFusco, Nathanson, and Zwick (2020) were more prone to speculation during the 2000s period.

4.2 Sensitivity to Capital Gains and Rental Yields

In this section we estimate the probability of a short-term sale given the potential capital gains and rental yield at the time of sale. We estimate logit models, like in Gilbukh and Goldsmith-Pinkham (2023). Table 3 contains summary statistics of the key variables we use for the logistic regressions. To estimate the sensitivity of investors to the potential capital gains, we estimate the following logit model for all purchases by retail investors, separate by each year of purchase:

$$l(Sell)_{i,j} = \beta_0 + \beta_1 G_j + C_j + C_i + C_c + u_{i,j}, \tag{1}$$

---

[9]We remove the sales that happen within the first 6 months out of an abundance of caution, to ensure that we are not capturing duplicate deeds for the same property. Including those transactions does not change the results.

where i indexes the investor (buyer name) and j the house (unique identifier). $l(Sell)_{i,j} = \log(\frac{\pi_{i,j}}{1-\pi_{i,j}})$, where $\pi_{i,j}$ is the probability that property j that was bought by investor i to be sold within 2 years. We also show the results for the probability that the property is sold in 3 years.

The variable G_j is the price growth that the house experiences the year after purchase, calculated as the price growth in the zip code of the house.[10] Using the price index of the median house in the zip code to calculate the potential capital gains has several advantages. First, we can assign capital gains to both the houses that were sold and the houses that were not sold within two years. Second, the Zillow Home Value Index allows us to have a comprehensive picture of the dynamics of prices, as it considers all homes in the calculation of the index, irrespective of when they were built or when they have been listed on the market. Third, we avoid using realized capital gains based on actual transaction prices that might depend on idiosyncratic features of the property or renovation cost. We use the price growth as the proxy for the potential capital gains an investor is facing when deciding to sell. Taxes, transaction fees, management costs, etc. are absorbed by the MSA fixed effects.

The specification includes an array of controls for the property, the investor and economic factors, which mitigate concerns of omitted variable bias. C_j is an array of property characteristics: actual purchase price from the deeds, log size in square feet, number of rooms, and age of the house or years from the last renovation (effective age) if available.[11] C_i is an array of investor characteristics: a dummy for local investors (their address is in the same MSA as the house), a dummy for foreign investors and a dummy for legal entities. C_c is an array of economic variables at the county-year level: population growth, income growth and unemployment rate change during the year before the sale of the house. These variables capture the dynamics of local housing demand. The specification also controls for MSA fixed effects to further control for demand factors. It also controls for the month of purchase to account for seasonality effects. The estimation allows for robust standard errors, clustered by zipcode.

Table 4 shows the results of the estimations, and Figure 10 plots the estimated coefficients and 95% confidence intervals. The investors pre-crisis have high sensitivity to the price growth in their decision to sell in the short term. The coefficients β_1 are 0.13 or above for the decision to sell in 2 or 3 years, for purchases in 2001-04. Those coefficients become close to 0.03 or statistically indistinguishable from zero, for purchases in 2012-15. This result shows evidence that the probability to sell in the short-term is four to six times less sensitive, if at all, to the

---

[10]Alternatively, we calculate G_j as the price growth in the zip code of the house in the two years after purchase, or the annual price growth beginning one year after purchase.

[11]Including as a control only the actual age of the house, or separately the actual age and effective age does not change the results.

price growth post-crisis compared to pre-crisis.

To study the sensitivity of investors to the rental yield, we estimate the following logit model:

$$l(Sell)_{i,j} = \beta_0 + \beta_1 R_j + C_j + C_i + C_c + u_{i,j}. \tag{2}$$

The dependent variable $l(Sell)_{i,j}$, the controls C_j, C_i, C_c, and the dummies for the month are like before. R_j is the rental yield that the house experiences the year after purchase, calculated as the average rent-to-price ratio in the MSA of the house.[12] The specification also controls for state fixed effects to further control for state-specific taxes, costs, and demand factors. We estimate the logit model for all purchases by retail investors, separate in each year. The estimation allows for robust standard errors, clustered by MSA.

Table 5 shows the results for sensitivities to rental yields, and Figure 11 plots the estimated coefficients and 95% confidence intervals. We find evidence that the higher the rental yield the lower the probability of sale in the short-term, as indicated by the negative coefficient β_1 throughout the period we study. Importantly, β_1 is becoming approximately double in absolute value in the post-crisis years, which is evidence that the investors are becoming more sensitive to the rental yield.

In the next part of our analysis we include all years together. We estimate the probability of selling an investment property based on the price growth and rental yield at the location of the property, interacted with a dummy for the post-crisis period. Since the dataset is enormous, to save on estimation time we use a random sample of 50% of the transactions in each year. The logit model is as follows:

$$l(Sell)_{i,j,t} = \beta_0 + \beta_1 G_{j,t} Post_t + \beta_2 G_{j,t} + C_j + C_i + C_{c,t} + u_{i,j,t}, \tag{3}$$

where i indexes the investor, j the house and t the date of purchase. $l(Sell)_{i,j,t} = \log(\frac{\pi_{i,j,t}}{1-\pi_{i,j,t}})$, where $\pi_{i,j,t}$ is the probability of the property j that was bought by investor i on date t to be sold within 2 years or 3 years. $G_{j,t}$ is the price growth that the house experiences the year after purchase, calculated as the price growth in the zip code of the house.[13] $Post_t$ is a dummy that takes the value of zero in the pre-crisis boom 2001-04 and the value of 1 in the post-crisis boom 2012-15. The arrays of controls C_j, C_i, $C_{c,t}$ and month dummies are like before. In this

---

[12]Alternatively, we calculate R_j as the rent-to-price ratio in the MSA of the house during the two years after purchase, or the rent-to-price ratio during the year that begins one year after purchase.

[13]Alternatively, we calculate $G_{j,t}$ as the price growth in the zip code of the house in the two years after purchase, or the annual price growth beginning one year after purchase.

specification we control additionally for MSA and year fixed effects or MSA×year fixed effects. The estimation allows for robust standard errors, clustered by zip code.

Table 6 shows the results of the estimation of equation (3). Again, we find evidence that the new investors have lower sensitivity to past price growth, in their decision to sell their investments in the short-term. The relation of price growth and probability of sale is positive and significant overall, but the interaction of the price growth with the $Post_t$ dummy has a negative and significant relation with the probability of sale. Figure 12 plots the estimated probability of sale within 2 years for a range of values of price growth, using the specification with MSA×year fixed effects. Pre-crisis there is a positive slope for the probability of short-term sale as potential capital gains increase, but this slope becomes close to zero post-crisis. Specifically, for one standard deviation higher house price appreciation from the mean, the probability of selling within two years is 4.0% higher for pre-crisis investors and 0.3% higher for post-crisis investors.

The specification of the logit model for rental yield is as follows:

$$l(Sell)_{i,j,t} = \beta_0 + \beta_1 R_{j,t} Post_t + \beta_2 R_{j,t} + C_j + C_i + C_{c,t} + u_{i,j,t}. \tag{4}$$

$R_{j,t}$ is the rental yield of the house the year after purchase, calculated as the average rent-to-price ratio in the MSA of the house.[14] The other variables are like the previous equation. We control additionally for state and year fixed effects or state×year fixed effects. The estimation allows for robust standard errors, clustered by MSA.

Table 7 shows the results of the estimation of equation (4). The relation of rental yield and the probability of sale is negative and significant, and there is an additional negative relation in the post-crisis boom. Figure 13 plots the estimated probability of sale within 2 years for a range of values of rental yield, from the specification that includes state×year fixed effects. Pre-crisis there is a negative slope for the probability of short-term sale as rental yield increases, and this negative slope becomes steeper post-crisis. The economic significance of the results is substantial: For one standard deviation higher rental yield from the mean, the probability of selling within two years decreases by 1.8% for pre-crisis investors and by 8.2% for post-crisis investors.

The results in this section, point to pre-crisis speculators primarily reselling houses in a short period, without living in or renting them. Their motivation for the investment is capital

---

[14]Alternatively, we calculate $R_{j,t}$ as the rent-to-price ratio in the MSA of the house during the two years after purchase, or the rent-to-price ratio during the year that begins one year after purchase.

gains (Haughwout et al. 2011). In contrast, a buy-and-hold strategy is less reactive to capital gains, and investors hold properties longer in search for rental income. Our robust evidence, using a variety of specifications and variable definitions, shows a buy-and-hold strategy during the boom of the 2010s.

5   Implications of the New Facts

The previous findings are consistent with search for yield in the low-interest-rate environment that followed the Global Financial Crisis. Real estate was a relatively high-yielding asset class in such environment. Thus, as returns on safe assets fell in the post-crisis period, many wealthy households decided to become real estate investors. This aligns with the evidence in Daniel, Garlappi, and Xiao (2021) that investors in a low-interest-rate world seek out high-dividend stocks to place their savings. Martinez-Miera and Repullo (2017) provide a model of search-for-yield.

Our findings have important implications. Firstly, the decreased sensitivity of post-crisis investors to changes in mortgage rates due to reduced leverage may alter the traditional channels of monetary policy. This adjustment could potentially impact the effectiveness of central bank policies. Additionally, the reduced vulnerability of these investors to defaults in falling markets, compared to highly leveraged investors, has the potential to decrease the amplification of market shocks and mitigate systemic risks. This contributes to a more stable market environment during downturns.

Secondly, it is important to consider the behavioral implications of the change in the type of investors. According to the literature, wealthier, more educated, and generally more sophisticated investors exhibit fewer investment biases. Examples in this literature are Agnew (2006), Dhar and Zhu (2006), Kumar (2009), Calvet, Campbell, and Sodini (2009), Bose, MacDonald, and Tsoukas(2015), Cronqvist and Siegel (2015) and Gomes, Haliassos, and Ramadorai (2021), among others. Thus, our findings suggest that the new investors are less prone to behavioral biases and irrational exuberance, potentially reducing the likelihood of another housing market collapse akin to the late 2000s.

Another important implication of the new facts pertains to the liquidity of housing markets. Liquidity is expected to decrease as the new investors reduce turnover and housing supply. Buy-and-hold investors engage in fewer transactions. Therefore, post-crisis we anticipate less inventory for sale.

Finally, our findings have implications that extend to the current economic environment. The long investment horizons and preferences for the inflation-protected cash flows that housing provides could explain why the housing market has been extremely insensitive to the tightening cycle that the Federal Reserve started in 2022.

6   Conclusions

This paper undertakes a thorough data exercise and shows a radical change in the composition of retail investors in housing markets after the Global Financial Crisis. New households entered the market. Post-GFC investors have several crucial distinctions: longer holding periods, reduced sensitivity to short-term capital gains, increased reliance on rental yields, greater wealth and education, creation of more legal entities, and a preference for cash over mortgage financing.

Notably, our findings indicate a distinct change in investment behavior, likely influenced by a combination of factors, including monetary policy and the pursuit of alternative yield-generating assets. As we navigate an evolving economic landscape marked by inflationary pressures and changing asset class correlations, the role of housing as a hedge against inflation gains prominence, potentially shaping the behavior of these investors in the future.

Our study unveils several unexplored avenues for future research. It reveals that during the Quantitative Easing period in the early 2010s, a new type of investors went into housing, which is an illiquid and lumpy asset. This fact has implications for the dynamics of household portfolio choice and retirement savings. Additionally, our new facts show a change in the ownership of the housing stock that becomes increasingly concentrated among wealthy households. The drop in homeownership following the Global Financial Crisis is likely to be persistent if the new investors are long-term. This will have implications for inequality and labor mobility.

In conclusion, our comprehensive analysis contributes to the evolving literature on housing market investors, offering novel insights into the characteristics, preferences, and implications of post-GFC investors. Our study informs policymakers, market participants, and researchers aiming to understand the dynamics of housing markets in today's economic context.

References

Agnew, J. R.: 2006, Do behavioral biases vary across individuals? Evidence from individual level 401(k) data, *Journal of financial and Quantitative Analysis* **41**(4), 939–962.

Albanesi, S., DeGiorgi, G. and Nosal, J.: 2022, Credit growth and the financial crisis: A new narrative, *Journal of Monetary Economics* **132**, 118–139.

Allen, M. T., Rutherford, J., Rutherford, R. and Yavas, A.: 2018, Impact of investors in distressed housing markets, *The Journal of Real Estate Finance and Economics* **56**(4), 622–652.

Amherst Capital: 2018, An update on institutional single-family rental activity. https://www.amherst.com/insights/an-update-on-institutional-single-family-rental-activity-2017-2018-u-s-market-trends-support-long-term-growth-opportunity/.

Bayer, P., Geissler, C., Mangum, K. and Roberts, J. W.: 2020, Speculators and middlemen: The strategy and performance of investors in the housing market, *The Review of Financial Studies* **33**(11), 5212–5247.

Bhutta, N.: 2015, The ins and outs of mortgage debt during the housing boom and bust, *Journal of Monetary Economics* **76**, 284–298.

Boddin, D., te Kaat, D., Ma, C. and Rebucci, A.: 2023, A housing portfolio channel of QE transmission.

Bose, U., MacDonald, R. and Tsoukas, S.: 2015, Education and the local equity bias around the world, *Journal of International Financial Markets, Institutions and Money* **39**, 65–88.

Calvet, L. E., Campbell, J. Y. and Sodini, P.: 2009, Measuring the financial sophistication of households, *American Economic Review* **99**(2), 393–398.

Chinco, A. and Mayer, C.: 2016, Misinformed speculators and mispricing in the housing market, *The Review of Financial Studies* **29**(2), 486–522.

Cronqvist, H. and Siegel, S.: 2014, The genetics of investment biases, *Journal of Financial Economics* **113**(2), 215–234.

Cvijanović, D. and Spaenjers, C.: 2021, 'We'll always have Paris': Out-of-country buyers in the housing market, *Management Science* **67**(7), 4120–4138.

Daniel, K., Garlappi, L. and Xiao, K.: 2021, Monetary policy and reaching for income, *The Journal of Finance* **76**(3), 1145–1193.

Davids, A. and Georg, C. P.: 2020, The cape of good homes: Exchange rate depreciations, foreign demand and house prices.

DeFusco, A. A., Nathanson, C. G. and Zwick, E.: 2022, Speculative dynamics of prices and volume, *Journal of Financial Economics* **146**(1), 205–229.

Demers, A. and Eisfeldt, A. L.: 2022, Total returns to single-family rentals, *Real Estate Economics* **50**(1), 7–32.

Dhar, R. and Zhu, N.: 2006, Up close and personal: Investor sophistication and the disposition effect, *Management Science* **52**(5), 726–740.

Eichholtz, P., Korevaar, M., Lindenthal, T. and Tallec, R.: 2021, The total return and risk to residential real estate, *The Review of Financial Studies* **34**(8), 3608–3646.

Favilukis, J. and Van Nieuwerburgh, S.: 2021, Out-of-town home buyers and city welfare, *The Journal of Finance* **76**(5), 2577–2638.

Gao, Z., Sockin, M. and Xiong, W.: 2020, Economic consequences of housing speculation, *The Review of Financial Studies* **33**(11), 5248–5287.

Garcia, D.: 2022, Second-home buying and the housing boom and bust, *Real Estate Economics* **50**(1), 33–58.

Garriga, C., Gete, P. and Tsouderou, A.: 2023, The economic effects of real estate investors, *Real Estate Economics* **51**(3), 655–685.

Gete, P. and Reher, M.: 2016, Two extensive margins of credit and loan-to-value policies, *Journal of Money, Credit and Banking* **48**(7), 1397–1438.

Gilbukh, S. and Goldsmith-Pinkham, P.: 2023, Heterogeneous real estate agents and the housing cycle. National Bureau of Economic Research Working Paper 31683.

Glaeser, E. L. and Nathanson, C. G.: 2017, An extrapolative model of house price dynamics, *Journal of Financial Economics* **126**(1), 147–170.

Gomes, F., Haliassos, M. and Ramadorai, T.: 2021, Household finance, *Journal of Economic Literature* **59**(3), 919–1000.

Gurun, U. G., Wu, J., Xiao, S. C. and Xiao, S. W.: 2023, Do Wall Street landlords undermine renters' welfare?, *The Review of Financial Studies* **36**(1), 70–121.

Haughwout, A., Lee, D., Tracy, J. S. and Van der Klaauw, W.: 2011, Real estate investors, the leverage cycle, and the housing market crisis. Staff Report, No. 514. Federal Reserve Bank of New York, NY, USA.

Korevaar, M.: 2023, Reaching for yield and the housing market: Evidence from 18th-century Amsterdam, *Journal of Financial Economics* **148**(3), 273–296.

Kumar, A.: 2009, Who gambles in the stock market?, *The Journal of Finance* **64**(4), 1889–1933.

Lambie-Hanson, L., Li, W. and Slonkosky, M.: 2022, Real estate investors and the U.S. housing recovery, *Real Estate Economics* **50**(6), 1425–1461.

Loewenstein, L. and Willen, P. S.: 2023, House prices and rents in the 21st century. National Bureau of Economic Research Working Paper 31013.

Martinez-Miera, D. and Repullo, R.: 2017, Search for yield, *Econometrica* **85**(2), 351–378.

Mills, J., Molloy, R. and Zarutskie, R.: 2019, Large-scale buy-to-rent investors in the single-family housing market: The emergence of a new asset class, *Real Estate Economics* **47**(2), 399–430.

Piazzesi, M. and Schneider, M.: 2009, Momentum traders in the housing market: Survey evidence and a search model, *American Economic Review* **99**(2), 406–11.

Sufi, A. and Taylor, A. M.: 2022, Financial crises: A survey, *Handbook of International Economics* **6**, 291–340.
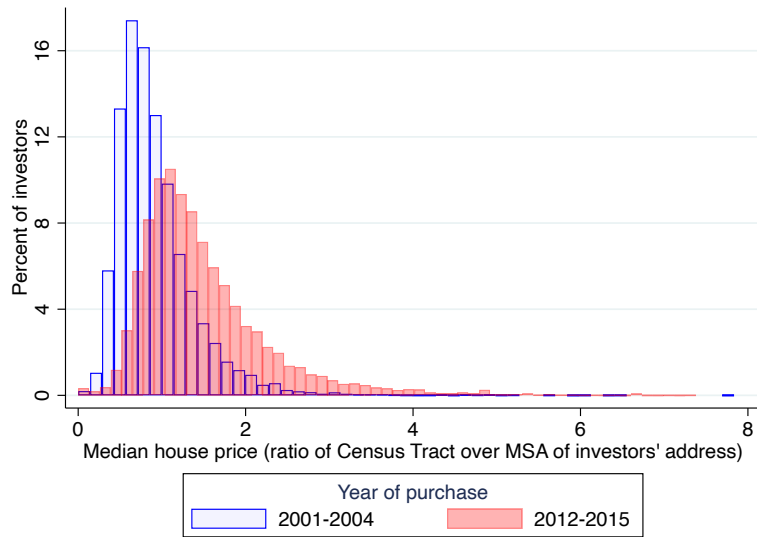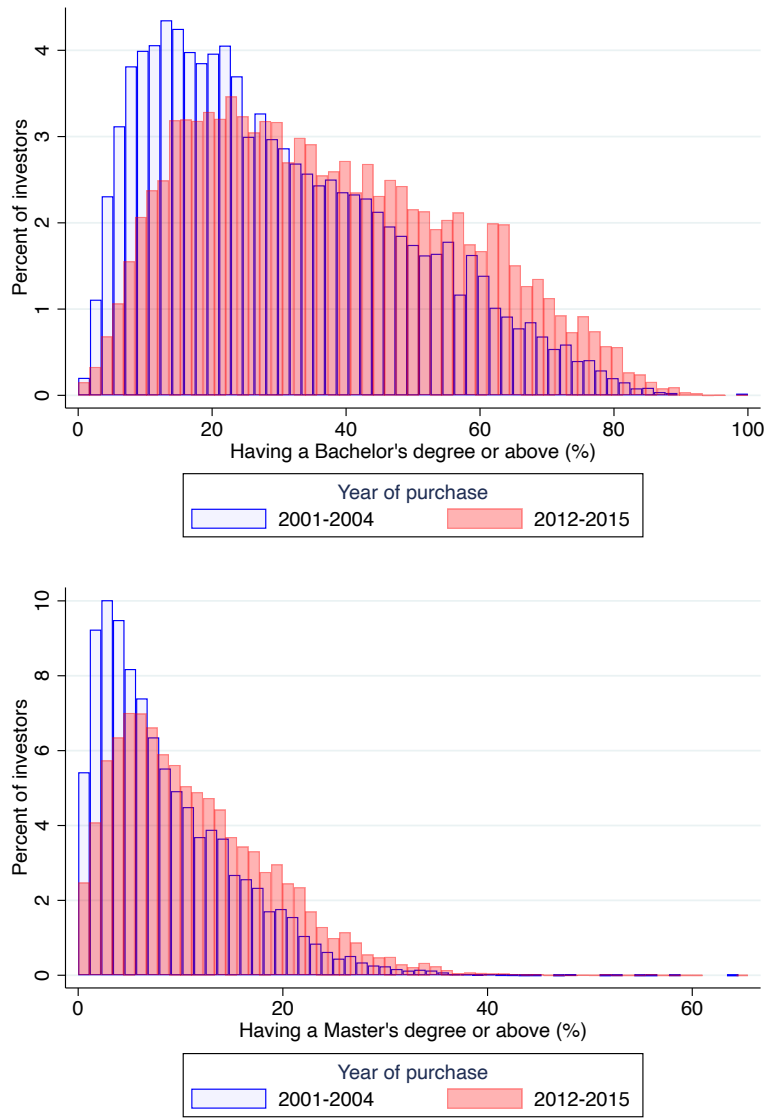
Figures
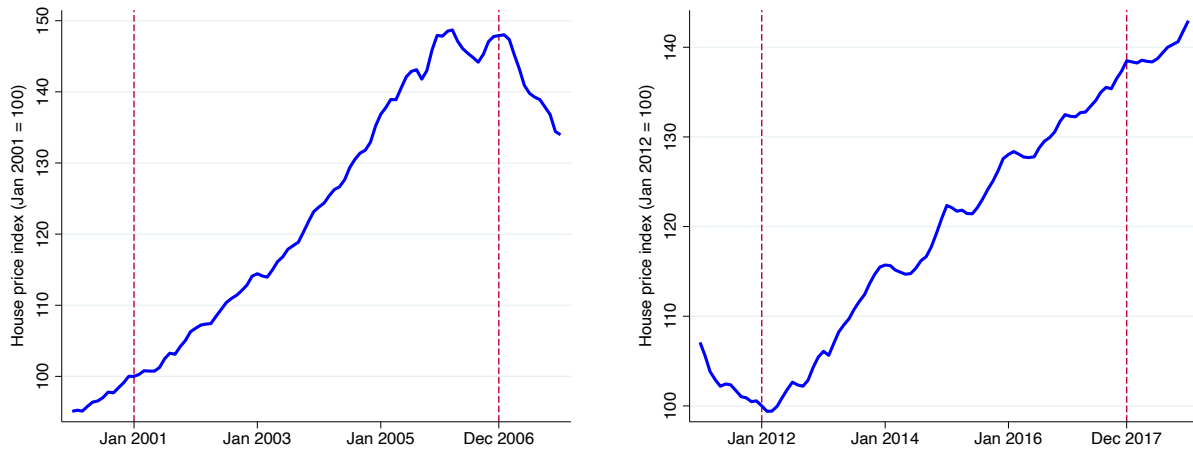
Figure 1. Two housing booms. The left panel plots the monthly inflation-adjusted S&P/Case-Shiller U.S. National Home Price Index during the years 2000-07. The index is calibrated to be 100 in January 2001. The right panel plots this index for the years 2011-18, calibrated to be 100 in January 2012. To adjust for inflation we use the Consumer Price Index for All Urban Consumers: All Items Less Shelter in U.S. City Average. The dashed vertical lines mark the two 6-year periods of housing boom we focus on. The time series come from FRED.
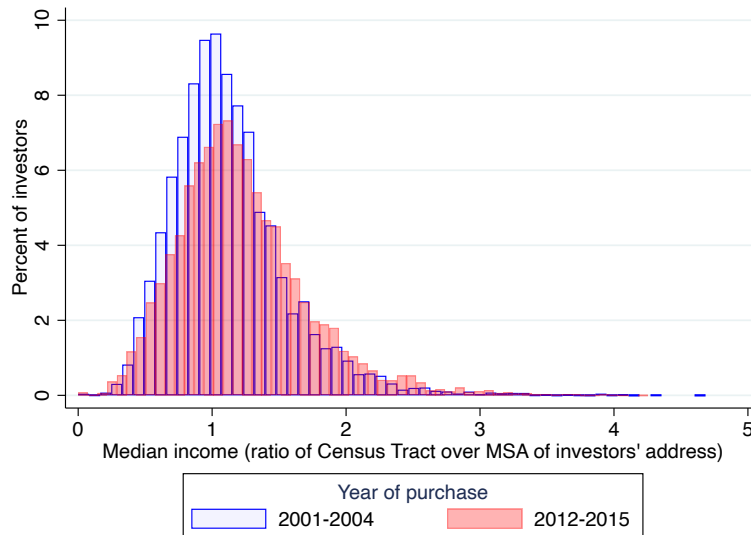
Figure 2. Distribution of investors' income. The histogram shows the distribution of the ratio of the census tract median family income over the MSA median family income where the investors' address is located. Ratio of 1 means the census tract median income equals the MSA median income. Ratios below (above) 1 mean that the census tract has lower (higher) median income than the MSA it belongs to.

Figure 3. Distribution of investors' house value. The histogram shows the distribution of the ratio of the census tract median house price over the MSA median house price where the investors' address is located.

Figure 4. Distribution of investors' education level. The top histogram shows the distribution of the share of people above 25 years old with highest level of education a Bachelor's, a Master's, a Doctorate or a Professional Degree in the census tract where the investors' address is located. The bottom histogram shows the distribution of the share of people above 25 years old with highest level of education a Master's or a Doctorate Degree in the census tract where the investors' address is located.
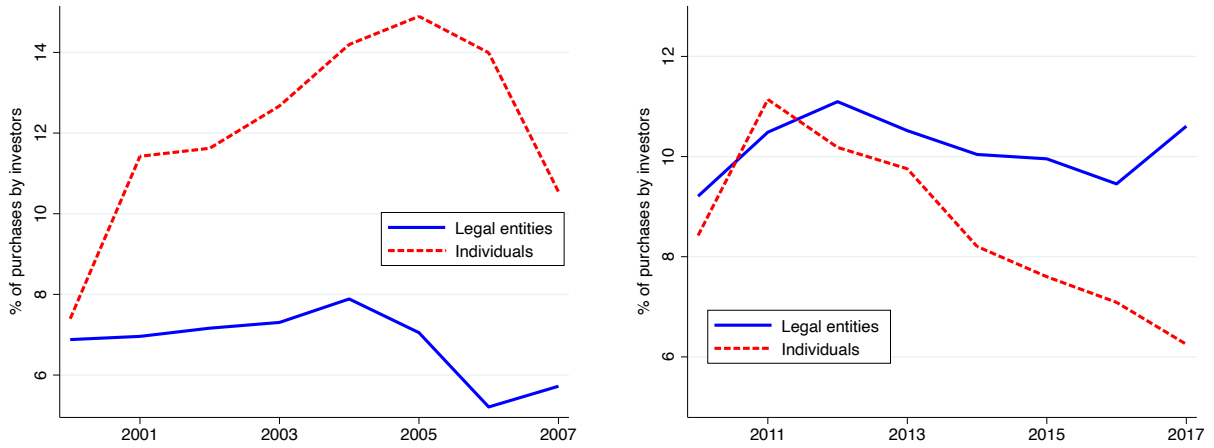
Figure 5. Retail investors' sophistication. The left panel plots the percentage of total number of purchases of single-family homes by retail investors (flow of investment) during the years 2000-07. The solid line plots the percentage of legal entities and the dashed line the percentage of individual investors. The right panel plots the flow of investment from these two types of investors during the years 2010-17. Legal entities are small or medium-sized buyers who purchase homes through an LLC, LP, Trust etc. Individual investors are individuals who purchase two or more houses within two years in the same MSA.
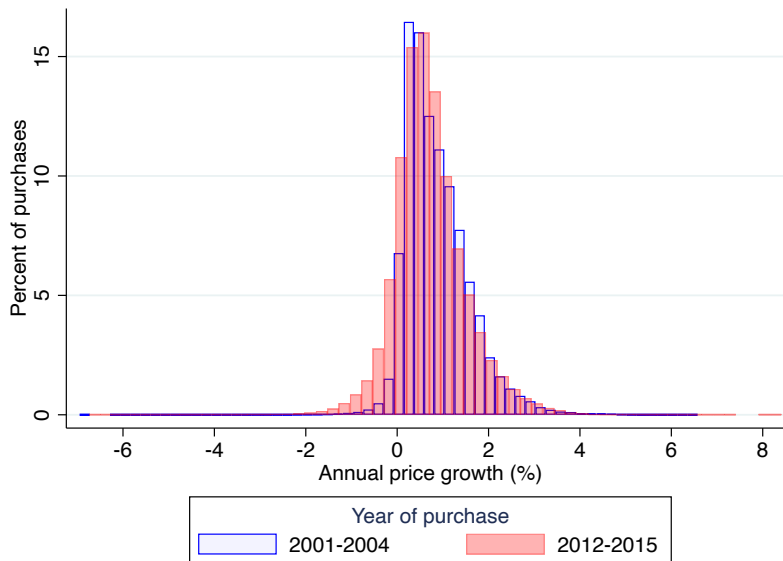
Figure 6. Distribution of house price growth in the areas of investment properties. The histogram shows the distribution of the annual price growth across the zip codes in which the investment properties are located. The histograms compare these distributions for properties purchased in 2001-2004 versus properties purchased in 2012-2015.
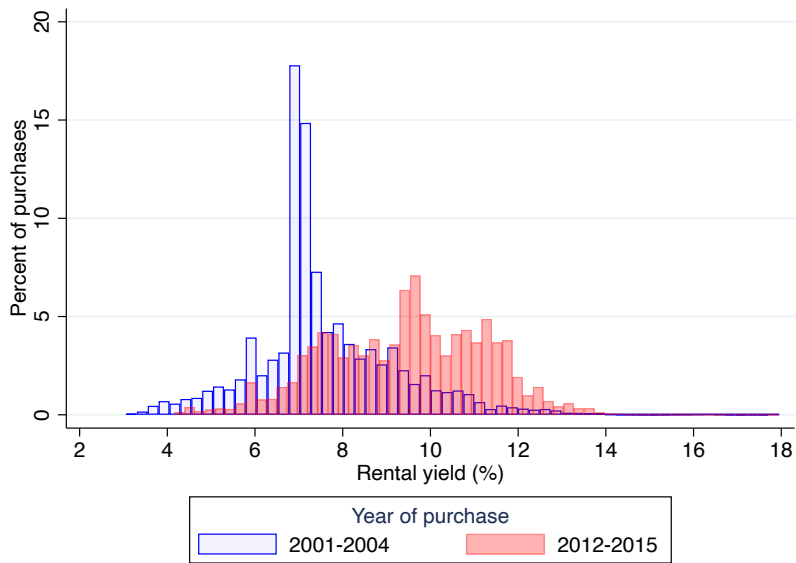
Figure 7. Distribution of rental yield in the areas of investment properties. The histograms show the distribution of the rental yield across the MSAs in which the investment properties are located. The histograms compare these distributions for properties purchased in 2001-2004 versus properties purchased in 2012-2015.
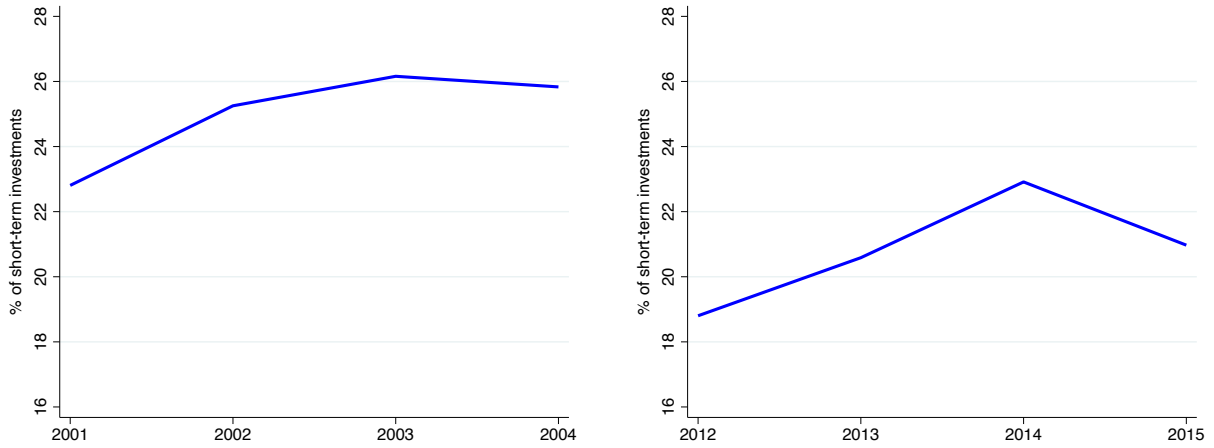
Figure 8. Short-term sales by investors. The figures plot the percentage of houses bought by retail investors in a particular year that were subsequently sold between 6 months and 3 years of their purchase. The left panel plots this percentage for the houses bought during the pre-GFC boom, 2001-04, and the right panel during the post-GFC boom, 2012-15. We calculate the holding period as the number of days from the date the deed was signed to buy a particular house, until the date the following deed was signed to sell the house.
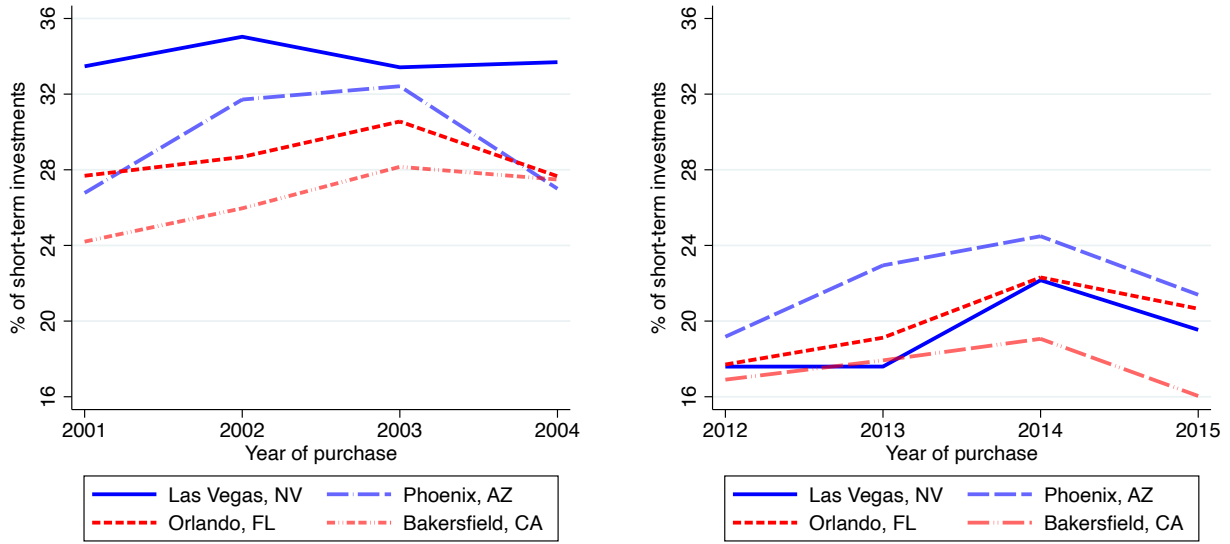
Figure 9. Short-term sales by investors in MSAs with the largest boom–bust cycles during the 2000s. The figures plot the percentage of houses bought by retail investors in a particular year that were subsequently sold between 6 months and 3 years of their purchase. The left panel plots this percentage for the houses bought during the pre-GFC boom, 2001-04, and the right panel during the post-GFC boom, 2012-15. We calculate the holding period as the number of days from the date the deed was signed to buy a particular house, until the date the following deed was signed to sell the house.
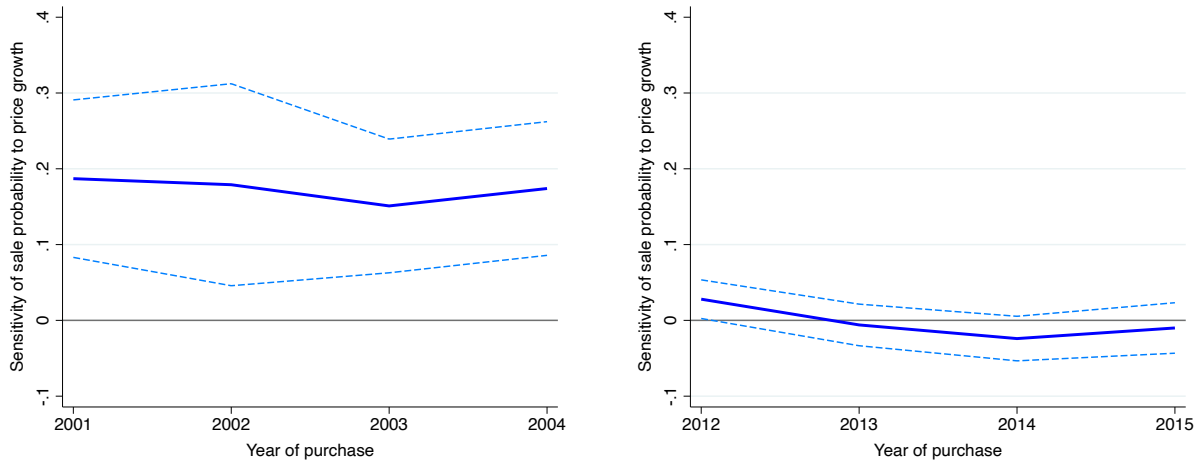
Figure 10. Sensitivity of probability of short-term sale to house price growth.
The figure plots the estimated coefficients and the 95% confidence intervals of the regression of the probability of sale in two years from the date of purchase on the house price growth for each year of purchase. The coefficients are derived from the estimation of the logit model (1) described in Section 4.
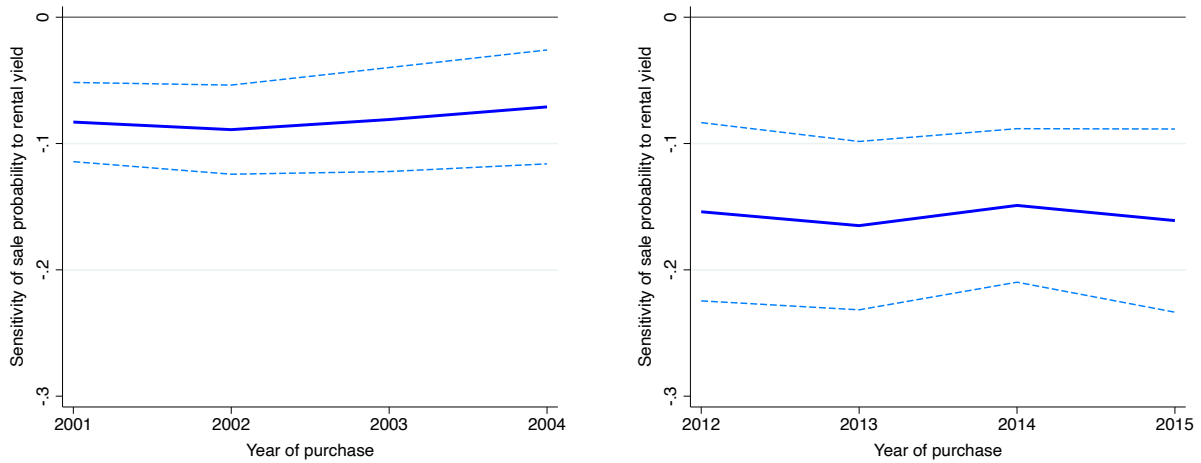
Figure 11. Sensitivity of probability of short-term sale to rental yield. The figure plots the estimated coefficients and the 95% confidence intervals of the regression of the probability of sale in two years from the date of purchase on the rental yield for each year of purchase. The coefficients are derived from the estimation of the logit model (2) described in Section 4.
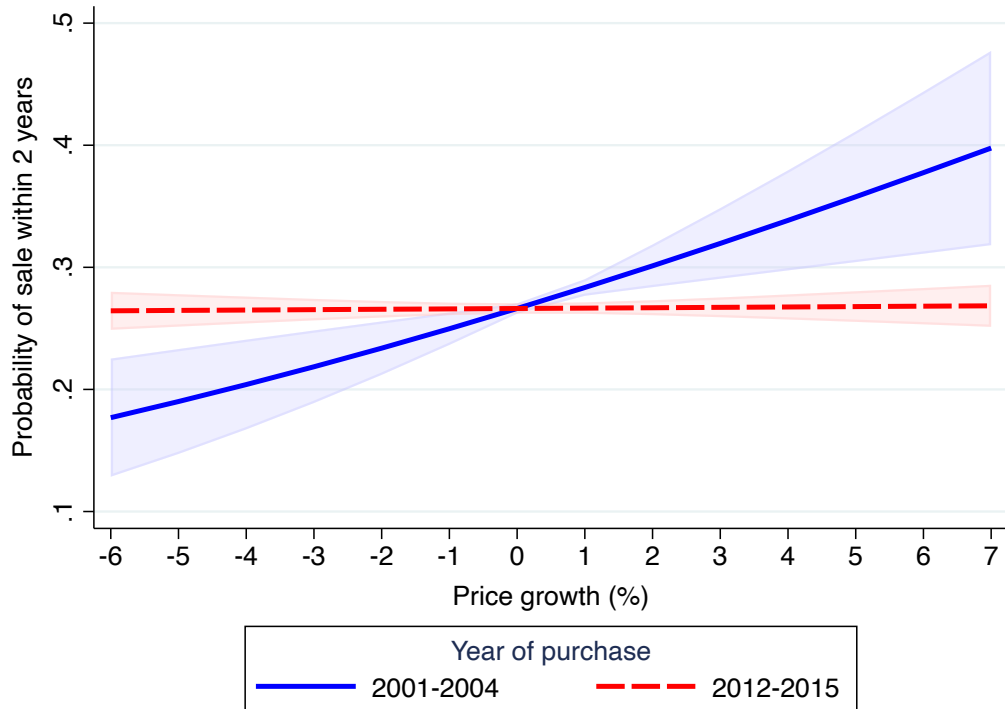
Figure 12. Probability of short-term sale and house price growth pre- and post-crisis. The figure plots the estimated probability of sale within two years from the date of purchase for different values of house price growth. The probabilities are shown for the properties purchased by investors pre-crisis between 2001 and 2004, and post-crisis between 2012 and 2015. The shaded areas show the 95% confidence intervals. The probabilities are derived from the estimation of the logit model (3) described in Section 4.
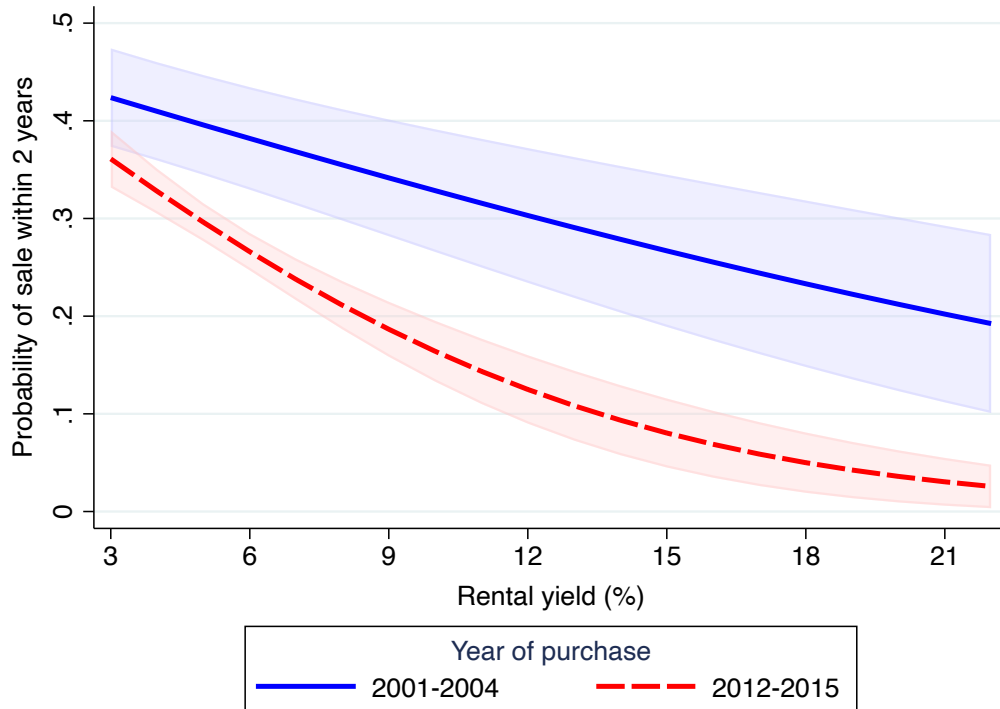
Figure 13. Probability of sale in two years and rental yield pre- and post-crisis.
The figure plots the estimated probability of sale within two years from the date of purchase
for different values of rental yield. The probabilities are shown for the properties purchased by
investors pre-crisis between 2001 and 2004, and post-crisis between 2012 and 2015. The shaded
areas show the 95% confidence intervals. The probabilities are derived from the estimation of
the logit model (4) described in Section 4.

Tables

Table 1. Comparing investors in two booms

	Pre-GFC boom (2001-2004)		Post-GFC boom (2012-2015)		Comparison
	Mean	SD	Mean	SD	Difference
Median income ($000)	54.4	23.4	69.1	30.8	14.7***
Income ratio, census tract to MSA	1.12	0.44	1.22	0.49	0.09***
Owner-occupied house price ($000)	158	101	284	186	126***
House price ratio, census tract to MSA	0.92	0.48	1.50	0.80	0.58***
Bachelor's degree or above (%)	30.2	18.62	37.2	19.6	7.01***
Master's degree or above (%)	8.66	6.73	11.4	7.58	2.69***
Legal entity (%)	39.5	48.9	58.8	49.2	19.2***
Observations	2,264,664		1,963,712		

This table summarizes the characteristics of retail investors who bought homes during 2001-04, and 2012-15, and shows the results of t-tests on the equality of the means. Each observation is a house purchase by a retail investor. ***sig. at 1%

Table 2. Investors' leverage per year

Pre-GFC boom			Post-GFC boom		
Year	Has mortgage (%)	Loan-to-value ratio	Year	Has mortgage (%)	Loan-to-value ratio
2001	64.55	0.849	2012	19.69	0.810
2002	64.14	0.842	2013	26.54	0.807
2003	64.38	0.833	2014	23.01	0.811
2004	65.37	0.817	2015	22.73	0.821
Average	64.68	0.833	2015	23.06	0.812

This table shows for each year, the share of properties bought by investors with mortgage and the average loan-to-value ratio, conditional on having a mortgage. The difference in the share of investors with mortgage is -41.62pp (p<0.01) and in average loan-to-value ratio is -0.021 (p<0.01), based on t-tests on the equality of means.

Table 3. Variables for sensitivity to capital gains and rental yields analysis

	Pre-GFC boom (2001-2004)				Post-GFC boom (2012-2015)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Sale within two years	0.17	0.38	0.00	1.00	0.16	0.37	0.00	1.00
Sale within three years	0.25	0.43	0.00	1.00	0.20	0.40	0.00	1.00
Price growth (%)	0.87	0.70	-7.22	6.58	0.73	0.79	-6.75	8.41
Rental yield (%)	7.59	1.70	3.06	22.8	9.43	1.82	4.16	20.6
Sales price ($thousand)	204	161	13	1,216	212	243	12	2,199
House age (years)	30.5	29.2	0.00	109	42.5	30.1	0.00	122
House log size	9.09	0.94	0.00	13.0	9.09	0.96	0.00	13.4
Number of rooms	3.01	3.37	0.00	10.0	2.72	3.23	0.00	10.0
Local buyer	0.63	0.48	0.00	1.00	0.70	0.46	0.00	1.00
Foreign buyer	0.001	0.03	0.00	1.00	0.01	0.09	0.00	1.00
Legal entity	0.34	0.47	0.00	1.00	0.55	0.50	0.00	1.00
Population growth (%)	1.74	1.83	-2.12	10.2	1.04	0.88	-5.46	8.99
Unempl. rate change (pp)	0.24	0.74	-2.00	3.20	-1.05	0.53	-2.90	2.30
Income growth (%)	5.20	3.74	-11.1	44.9	4.31	3.16	-13.8	21.7
Observations	1,557,752				1,387,427			

This table presents summary statistics of the key variables at the property level. The sample consists of the universe of single-family home purchases by retail investors that we could merge with prices, rents, property characteristics, and economic controls. We report the summary statistics for the house purchases in the years 2001-04, and 2012-15, which we track until their next sale, up to the end of 2017. The price growth is at the zip code of the property and the rental yield is at the MSA of the property. Population growth, unemployment rate change and per capita income growth are annual percentages at the county level.

Table 4. Probability of short-term sale and capital gains by year

	Pre-GFC				Post-GFC			
Year of purchase	2001	2002	2003	2004	2012	2013	2014	2015
	Sale within 2 years				Sale within 2 years			
Price growth	0.187***	0.179***	0.151***	0.174***	0.028**	-0.006	-0.024	-0.010
	(0.053)	(0.068)	(0.045)	(0.045)	(0.013)	(0.014)	(0.015)	(0.017)
	Sale within 3 years				Sale within 3 years			
Price growth	0.147***	0.154**	0.128***	0.156***	0.035***	0.005	-0.018	0.002
	(0.052)	(0.066)	(0.041)	(0.043)	(0.012)	(0.013)	(0.014)	(0.016)
Property controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investor controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demand controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	203,335	235,618	271,040	340,031	233,494	237,328	214,220	220,864

Robust standard errors clustered by zip code are in parentheses. This table presents the results from logistic regressions based on equation (1) and estimated for each purchase year separately. The top panel shows the results for the probability of selling the property within two years of purchase, calculated based on the exact date of purchase. The second panel shows the results for the probability of selling the property within three years of purchase. The property characteristics controls are the purchase price from the deed, age of the house, log size of the house and number of rooms. The investor type controls are dummies for local investors, foreign investors and legal entities. The controls for local demand are the population growth, income growth, unemployment rate change, and dummies for the month of purchase. All models include MSA fixed effects. Figure 10 plots the coefficients. ***sig. at 1%; **sig. at 5%.

Table 5. Probability of short-term sale and rental yield by year

Year of purchase	Pre-GFC				Post-GFC			
	2001	2002	2003	2004	2012	2013	2014	2015
	Sale within 2 years				Sale within 2 years			
Rental yield	-0.083***	-0.089***	-0.081***	-0.071***	-0.154***	-0.165***	-0.149***	-0.161***
	(0.016)	(0.018)	(0.021)	(0.023)	(0.036)	(0.034)	(0.031)	(0.037)
	Sale within 3 years				Sale within 3 years			
Rental yield	-0.066***	-0.068***	-0.063***	-0.070***	-0.128***	-0.140***	-0.123***	-0.152***
	(0.014)	(0.015)	(0.020)	(0.022)	(0.032)	(0.030)	(0.028)	(0.034)
Property controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investor controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demand controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	203,335	235,618	271,040	340,031	233,494	237,328	214,220	220,864

Robust standard errors clustered by MSA are in parentheses. This table presents the results from logistic regressions based on equation (2) and estimated for each purchase year separately. The top panel shows the results for the probability of selling the property within two years of purchase, calculated based on the exact date of purchase. The second panel shows the results for the probability of selling the property within three years of purchase. The controls for property characteristics, investor type and local demand are like in Table 4. All models include state fixed effects. Figure 11 plots the coefficients. ***sig. at 1%.

Table 6. Probability of short-term sale, price growth and retail investors

	Sale within 2 years		Sale within 3 years	
Price growth × Post-GFC investors	-0.052***	-0.095***	-0.035***	-0.072**
	(0.020)	(0.030)	(0.009)	(0.030)
Price growth	0.056***	0.097***	0.040**	0.081***
	(0.019)	(0.029)	(0.017)	(0.028)
Property characteristics	Yes	Yes	Yes	Yes
Investor type	Yes	Yes	Yes	Yes
Demand factors	Yes	Yes	Yes	Yes
MSA fixed effects	Yes	No	Yes	No
Year fixed effects	Yes	No	Yes	No
MSA × Year fixed effects	No	Yes	No	Yes
Observations	1,283,070	1,282,649	1,283,085	1,282,757

Robust standard errors clustered by zip code are in parentheses. The property characteristics controls are the purchase price on the deed, age of the house, log size of the house and number of rooms. The investor type controls are dummies for local investors, foreign investors and legal entities. The controls for local demand are the average growth in population, growth in per-capita income, and change in unemployment rate in the year before the potential sale, as well as month dummies for seasonality. The models include either MSA and year fixed effects or MSA×year fixed effects. ***sig. at 1%; **sig. at 5%.

Table 7. Probability of short-term sale, rental yield and retail investors

	Sale within 2 years		Sale within 3 years	
Rental yield × Post-GFC investors	-0.084***	-0.108***	-0.070***	-0.084***
	(0.028)	(0.035)	(0.024)	(0.031)
Rental yield	-0.055**	-0.072***	-0.043*	-0.065***
	(0.025)	(0.016)	(0.022)	(0.015)
Property characteristics	Yes	Yes	Yes	Yes
Investor type	Yes	Yes	Yes	Yes
Demand factors	Yes	Yes	Yes	Yes
State fixed effects	Yes	No	Yes	No
Year fixed effects	Yes	No	Yes	No
State × Year fixed effects	No	Yes	No	Yes
Observations	885,977	885,893	885,977	885,977

Robust standard errors clustered by MSA are in parentheses. The property characteristics controls are the purchase price on the deed, age of the house, log size of the house and number of rooms. The investor type controls are dummies for local investors, foreign investors and legal entities. The controls for local demand are the average growth in population, growth in per-capita income, and change in unemployment rate in the year before the potential sale, as well as month dummies for seasonality. The models include either state and year fixed effects or state×year fixed effects. ***sig. at 1%; **sig. at 5%; **sig. at 10%.

ONLINE APPENDIX (Not for Publication)

A   Detailed Description of Database

In this section we describe our data sources, how we cleaned the data, and the key variables we use in our analysis.

Investors' purchases

The transaction data come from the Deeds Dataset of Corelogic, a large database that collects all deeds for the transfer of ownership of single-family homes in the U.S. Each record contains the date of the transfer, the address of the property, the sale price, and the names of the buyer and seller. We keep transactions between January 1st, 2000 and December 31st, 2017.

We drop transactions with purchase price missing or smaller than $10,000, a common practice with deeds data. We also drop all transactions in the non-disclosure states. These states do not require that the sale price is submitted to the county office. Specifically the final dataset drops all records from the 12 non-disclosure states: Alaska, Idaho, Kansas, Louisiana, Mississippi, Missouri, Montana, New Mexico, North Dakota, Texas, Utah and Wyoming.

To identify legal entities we search for buyer names that contain the words "LLC," "LP," "INC," "TRUST," "CORPORATION," "PARTNERS," or entity names like "Invitation Homes". From these names we filter out names of relocation companies, non profit organizations, construction companies, national and regional authorities, banks, Ginnie Mae, Fannie Mae, Freddie Mac and other mortgage loan companies and credit unions, homeowner associations, names of the county, city or municipality. To identify relocation companies, non profit organizations and construction companies we use public data of lists of the top relocation companies, non profit organizations and construction companies in the U.S. We also manually check the names of the 200 largest non-individual buyers in each state using online search engines to classify them in the right category, and iterate this procedure several times to ensure the largest buyers are correctly classified.

To accurately classify the largest single-family institutional investors, which we exclude from the analysis, we collect from industry reports and news reports the names of the top 26 institutional investors in the single-family rental market. For example, Amherst Capital's 2018 market commentary report[15] provides a comprehensive list of the top 26 single-family

---

[15] Amherst Capital report is retrieved from https://www.amherst.com/insights/an-update-on-institutional-

rental institutions and the number of homes they own. We then search for the names of these top investors and their subsidiaries in the deeds database and ensure they are classified as institutional investors. We use public SEC filings and other business websites to track down the names of the subsidiaries of these large investors. This procedure results in calculating the exact holdings of the top single-family institutional investors.

To classify non-institutional buyers into individual retail investors we calculate the number of purchases of each individual name within the MSA within the given year and the year before. We define individuals retail investors as individuals who purchase more than one property within the MSA in the given year and the year before.

We calculate the share of retail investors as the number of purchases by either individual investors or small and medium-sized legal entities divided by the number of purchases of all purchases, by retail investors, large institutional investors and homeowners. The holding duration is the duration between the date the deed was signed for the purchase of the house and the date the subsequent deed was signed for the sale of the house.

We merge the deeds data with characteristics of the properties that come from tax assessors. The characteristics include the year the housing unit was built, total number of rooms, number of bedrooms, size of the house measured in square feet, and the year of the last renovation or update of the structure.

Demographic variables

We obtain data on the median income, education level, and owner-occupied house prices at the census tract from the Federal Financial Institutions Examination Council census and demographic data. We merge the census tract code with the census tract of the buyer mailing address in the deeds. This allows us to link each investor to the demographic and housing market data of the areas they live in.

House prices and rents

We match the property address in the deeds with data from various sources to obtain the evolution of housing prices and rents in the areas where investors go. In particular, house prices at the zip code level and monthly frequency come from the Zillow Home Value Index for single-family homes. This index is designed to capture the value of a typical property across

single-family-rental-activity-2017-2018-u-s-market-trends-support-long-term-growth-opportunity/

each zip code, not just the homes that sold, and it does so by utilizing information from the full distribution of homes in a given zip code. The housing rents also come from Zillow. The smallest geographical unit available for this series that goes back to the year 2000 is the MSA. We collect housing prices and rents with monthly frequency at the MSA level and calculate the rental yield as the ratio of 12 times the monthly rent over the price of the house.

Other variables

We also match the property address in the deeds with economic factors. We collect the following data at the county-year level, from 2000 to 2017: Population from the U.S. Census Bureau, per capita income from the from the Bureau of Economic Analysis, and unemployment rate from the U.S. Bureau of Labor Statistics. This allows us to link each property to the economic conditions of the county it is located in, at the time of purchase and at the time of potential sale, for example in two years from purchase.