The Economic Effects of Real Estate Investors*

Carlos Garriga†, Pedro Gete‡, Athena Tsouderou§

April 2022

Abstract

We show five new results concerning investors in U.S. housing markets. First, small investors have the largest growth across all cities post Great Recession. Second, large "Wall Street" investors concentrate in superstar cities. Third, the rise of small investors increases the price-to-income ratio, especially in the bottom price-tier. This effect is reversed as investors trigger a medium-run supply response. Fourth, in areas with a high supply elasticity, investors affect rents more than prices. Finally, investors change the composition of the housing stock in favor of multi-family units.

*We thank Itzhak Ben-David, Morris Davis, Anthony DeFusco, David Echeverry, Andra Ghent, Jonathan Halket, Lu Han, Nina Karnaukh, Finn Kydland, Jose Maria Liberti, David Ling, Christos Andreas Makridis, Charles Nathanson, Michael Reher, Stephen L. Ross, Martin Schneider, and participants at AREUEA, AREUEA International, Durham, ECB, Econometric Society, Hebrew University, HULM, NEOMA, Notre Dame, IE, Ohio State, SED, Spanish Finance Forum and Urban Economics Association. The views expressed herein do not necessarily reflect those of the Federal Reserve Bank of St. Louis or the Federal Reserve System. The results and opinions are those of the authors and do not reflect the position of Zillow Group. Research reported in this paper was partially funded by the Spanish Ministry of Economy and Competitiveness (MCIU), State Research Agency (AEI) and European Regional Development Fund (ERDF) Grant No. PGC2018-101745-A-I00. This paper was previously circulated as Investors and Housing Affordability.

†Email: carlos.garriga@stls.frb.org. Federal Reserve Bank of St. Louis, Research Division. P.O. Box 442, St. Louis, MO, 63166-0442. Phone: +1 (314) 444-7412.

‡Email: pedro.gete@ie.edu. IE Business School, IE University. Calle Maria de Molina 12, 28006 Madrid, Spain. Phone: +34 91 568 97 27.

§Email: athenatsouderou@gmail.com. Herbert Business School, University of Miami. 5250 University Dr, Coral Gables, FL 33146. Phone: +34 65 778 40 39.

1 Introduction

Real estate investors have recently attracted the attention of academia, policy and media circles. For example, officials in several cities have enacted or are discussing policies to block investors in housing markets.[1] In this paper we study real estate investors using a comprehensive database covering all U.S. housing transaction records for the period 2000-2017. First, we classify and document characteristics of real estate investors. Second, we use a novel identification strategy to study the effects of small investors on housing markets.

Concerning the characteristics of real estate investors, we show that small investors are the group with the largest growth in terms of housing purchases and value. These investors are mainly local as their purchases are in the same MSA of their mailing address. They are located throughout all the U.S. geography. The growth of these real estate investors takes place in a period characterized by a dramatic drop of risk-free rates. This is consistent with the portfolio channel documented by Daniel, Garlappi and Xiao (2021) for stock investments. Low and stable interest rates can lead to significantly higher demand for income-generating assets like housing.

Large investors, usually called "Wall Street Landlords," have been recently studied by several papers. See for example Allen et al. (2018), Mills, Molloy and Zarutskie (2019), Brunson and Buttimer (2020), Gurun et al. (2022), and Lambie-Hanson, Li and Slonkosky (2022). We show that contrary to small investors, large investors are geographically concentrated in large metropolitan areas, away from their mailing address.

To study the causal effects from investors, our identification strategy exploits a "Bartik inspired instrument" combining the Fed's Quantitative Easing (QE) programs and the fact that most investors are small and local. As yields from risk-free assets collapsed due to QE relatively high rental yields made local properties an attractive alternative investment. QE was a national shock and different regions reacted differently based on the pre-shock propensity for investments among the local high-income population. We capture such pre-shock "local propensity to invest" with the share of the top earners' business income over total income in each MSA in 2007. High income households that consistently receive business income are a great proxy for the number of sophisticated investors. The identification strategy is valid as long as after conditioning on multiple controls, the geographical distribution of top earners claiming business income in 2007 is uncorrelated with the factors that moved housing price

[1] For example New York and California, where the presence of investors reached unprecedented highs, approved statewide rent controls (Business Insider 2019). Amsterdam discussed banning investors from purchasing and renting properties (Bloomberg 2018), Berlin is considering expropriating large, private, profit-seeking landlords (The Wall Street Journal 2019), and Spain imposed measures to penalize investors (Bloomberg 2019).

dynamics between 2008-2017.

There are several reasons to trust the identification strategy: 1) We use many fixed effects and control variables that make it unlikely that the error term reflects common movers of both investors and housing market variables. Moreover, we exhaust the list of possible drivers of housing markets as controls: income, local economic activity, credit conditions, population, composition of labor markets, foreclosures... 2) The instrument seems random, it is very hard to predict it. Like in the entrepreneurship literature that shows that it is very hard to predict which cities become hubs for entrepreneurship (e.g. Davidsson 1991; Rocha and Sternberg 2005). Most of the cross-sectional differences are driven by random historical factors. Thus, most of the variation in the instrument seems unrelated to other drivers of housing markets. 3) A battery of tests suggest that the identification is valid. For example, areas with the highest or lowest levels of the instrument exhibit parallel pre-trends. Placebo tests confirm the parallel pre-trends.[2] Tests based on Altonji, Elder and Taber (2005) and Oster (2019) suggest no concerns of omitted variable bias. In addition, we show robustness to multiple alternative specifications and definitions of investors' purchases.

In terms of results, we show that one standard deviation higher purchases by investors leads to 1.46 percentage points higher housing price growth for the median house. Moreover, the market segment that is more sensitive to purchases by investors is the bottom price-tier. There are strong effects on the overall supply of housing with clear compositional effects in the characteristics of the newly constructed stock. One percentage point increase in the share of investors increases the number of new construction permits for single-unit buildings by 4.8% on average, and for buildings of 5 or more units by 16.4% on average.

To separate the short-run impact of investors (inelastic housing supply) from the long-run impact (housing supply can adjust) we apply the local projection method developed by Jordà (2005). The impact on prices weakens over time as new residential units are added to the stock of housing units. The effects also differ once we split the sample by the housing supply elasticity in each MSA based on Saiz (2010). In areas with a highly elastic supply of housing, the purchases of real estate investors affect more rents than house prices. Investors cause minimal price increases in MSAs where there are loose supply restrictions. In areas with a low housing supply elasticity, real estate investors have the opposite effect as prices increase more than rents.

Literature: The paper contributes to two literatures. The first one analyzes the recent emergence of corporate investors post-financial crisis. For example, Allen et al. (2018), Mills,

[2]This is strong support for the plausibility of the exogeneity assumption according to Goldsmith-Pinkham, Sorkin and Swift (2020).

Molloy and Zarutskie (2019) and Brunson and Buttimer (2020) describe this new class of investors. Ganduri, Xiao and Xiao (2019), Smith and Liu (2020) and Gurun et al. (2022) focus on large institutional investors. Graham (2020) studies implications of the investors during the housing bust of the 2000s, and Lambie-Hanson, Li and Slonkosky (2022) during the recovery from the Great Recession. Garriga, Gete and Tsouderou (2021) show that these investors are driven by search for yield. Moreover, Agarwal et al. (2019); Albanesi, De Giorgi and Nosal (2017), Bayer, Mangum and Roberts (2021) and Ben-David (2011) study short-term investors (commonly known as flippers). Chinco and Mayer (2016), Cvijanovic and Spaenjers (2021), Davids and Georg (2020) and Favilukis and Van Nieuwerburgh (2021) analyze foreign and out-of-town investors. We move forward this literature by highlighting the overwhelming increase in the small and local investors who buy through legal entities, and by studying their real effects.

The second contribution is to bring a finance perspective to the housing affordability literature. Traditionally, this literature emphasizes the role of housing supply constraints as a central issue leading to affordability problems (see for example Gyourko, Mayer and Sinai 2013, or Molloy, Nathanson and Paciorek 2022. Ben-Shahar, Gabriel and Oliner 2020 provide a survey). After classifying investors according to different types (i.e., by size, location, etc.), our analysis highlights that the purchases by some types that actively participate in real estate markets have a significant impact on house prices, rents, and hence affordability. The impact of investors interacts with housing supply elasticities, making the effect of purchases be large on prices in markets with low housing supply elasticity, and on rents in markets with high housing supply elasticity.

The rest of the paper is organized as follows: Section 2 describes the data and the characteristics of real estate investors. Section 3 presents the cross-sectional analysis. Section 4 presents the dynamic analysis. Section 5 assesses the validity of the instrument and the robustness of the results. Section 6 concludes.

2 Investors in Real Estate Markets

2.1 Data

The core data used in the analysis comes from the Zillow Transaction and Assessment Dataset (ZTRAX, Zillow 2017).[3]. The database covers all ownership transfers as recorded by the counties' deeds in the United States. The unit of account is based on individual ownership transfers

[3]We include a detailed description of the data sources in the Appendix *A*.

of residential properties, including multi-family and single-family, from January 1st, 2000 to December 31st, 2017. The sample period allows assessing the differences between the pre and post financial crises housing booms based on the participation of real estate investors. The final sample consists of about 85 million transactions.

The universe of deeds is characterized by buyers/owners of residential housing with different legal identities. We classify real estate investors based on the buyer name. Investors are legal entities who purchase homes using an LLC, LP, Trust, REIT, etc. in the purchase deed. We filter out from this category the buyers that are broadly defined as intermediaries, including relocation companies, non-profit organizations, construction companies and national and regional authorities, as well as banks, Ginnie Mae, Fannie Mae, Freddie Mac and other mortgage loan companies and credit unions, and the state taking ownership of foreclosed properties.

The key information used in the analysis relates to different measurements of the volume and share of purchases by all real estate investors. The key variable we construct measures the total dollar value of investors' purchases in real terms over all the purchases at the MSA-year level.[4]

To construct our instrumental variable we use zip code level information on the pre-crisis investment attitudes in different areas identified by individual tax returns from the Statistics of Income of the Internal Revenue Services (IRS). The data contains information of all the individuals filing income taxes by income group and by zip code, such as the number of individuals and total earnings. Our instrument is the average share of business income over total income of high earners (annual adjusted gross income above $100K) in each MSA in 2007. We weigh by the total income of high-earners to aggregate to the MSA level in 2007. The choice of the year is to specify their attitude towards investment before the financial crisis, and this avoids the share to move with house prices. As a robustness check, we have performed the analysis using previous years. To assess the shift of investments over time, we construct the panel version of the instrument by interacting with the average rate of one-year certificate of deposits (CD) rate from the consumer financial services company Bankrate.

For consistency with the transaction data, we use the Zillow Home Value indices for all homes, the bottom and the top-tier homes at the MSA level. The bottom-tier segment of the market is the bottom third of the housing price distribution in each MSA, and captures the typical rental unit that is attractive for real estate investors searching for cash-flow yield. The

[4]The number of purchases would underestimate presence in the apartment market. For example the number of purchases would equate a purchase of one condominium to the purchase of one apartment building of 100 apartments. For robustness checks we use alternative measures of the presence of investors based on the number of properties or the number of units purchased.

middle-tier captures the typical owner-occupied housing unit, whereas the top-tier captures the luxury market for owner and rental-occupied housing (i.e., including high-end vacation homes). For each group, the median price captures the average value within each segment (i.e., for the bottom tier the median price represents the 17th percentile of the prices of the total market).[5] Similarly, housing rents come from the Zillow Rent Index for all homes. For our comprehensive list of control variables we use population data from the Census, the unemployment rate from the Bureau of Labor Statistics, and income from the Statistics of Income of the IRS and Zillow. We calculate the 17th, 50th and 83rd percentiles of individual income from the IRS to get the price-to-income and the rent-to-price ratios for the corresponding tiers.

To explore the effects of investors' purchases in the supply side, we collect the number of new construction permits from the Census Bureau's annual Building Permits Survey available at the zip code level.

The data allows the inclusion of 332 MSAs with complete information on housing variables, investors' activity, control variables, and the instrumental variable. Table 1, Panel A summarizes the key statistics of the cross-sectional sample between 2009 and 2017. According to the data, investors purchase on average 12.37% of the market annually. The average house price growth (mid-tier) is 0.47% annually, and that includes some MSAs with house prices declines and other areas with nearly 6% growth annually. Table 1, Panel B summarizes the key variables in the panel analysis.

2.2 Trends of Real Estate Investors: A New Class of Investors

The trend of real estate purchases by investors for the period 2000-17 is summarized in Figure 1. As we can see, the share of housing purchases by investors dramatically increased after the Great Recession. Their participation changes right after the dramatic drop of risk-free rates (top panel) and ahead of the recovery of the stock market (bottom panel).[6]

We can further separate the real estate investors by their origins relative to the location of the investment. This creates three distinct groups with local, out-of-town domestic and foreign investors. Local investors have a mailing address in the purchase deed in the same MSA as the property purchased. Out-of-town domestic investors have their mailing address in the U.S., but outside the MSA of the property they purchase. Finally, foreign investors have a mailing

[5] In a symmetrical way, the top-tier segment of the market is the top third of the price distribution in each MSA, and the top-tier price is the top 83rd percentile of prices within an MSA.

[6] These patterns suggest a portfolio channel as what Daniel, Garlappi and Xiao (2021) show for stock investments.

address outside the U.S. Figure 2 plots the share of dollar purchases by these different categories of investors for the period 2000-17. After 2009, the purchases by local investors constituted more than 55% of the total purchases by investors. As documented in the next chart, local investors are more likely to be small in size, and the typical examples are mom-and-pop or business professionals that purchase homes in the MSA where they also live.

Out-of-town investors include the large REITs, having their headquarters and mailing address for example in New York, and purchasing properties in dense areas (i.e., Salisbury, MD) and vacationing areas (i.e. Naples, FL, Hilton Head Island, SC, San Rafael, CA). Out-of-town domestic investors account for about 35% of the purchases after 2009. For the transacted units with a complete address for the buyer, it appears that the market share for foreign investors is very small. This could be due to the fact that total foreign housing investment is not that large, or because foreign legal entities use a U.S. mailing address, in which case we classify them as domestic investors but most likely out-of-town investors.

According to Figure 1, in 2006 investors represented about 8% of the purchases whereas by 2015 they represented over 16% of the market. Where does the growth come from? Are these a large number of small investors or a small number of large ones? The transaction data allows us to answer the question by calculating the changes in the distribution of purchases by size and the number of investors by size between 2006 and 2015. Figure 3 shows the distribution of the total purchases of each investor by size (total real dollar value of purchases) in the years 2006 and 2015.[7] At the intensive margin (i.e. dollar amount) the top panel of Figure 4 shows that the small investors (i.e. below the 25th percentile of the size distribution), and, to a lesser extend, the very large investors (i.e. above the 95th percentile), had the largest growth in their purchases. The bottom panel of Figure 4 shows that at the extensive margin (i.e. number of investors) the increase is driven by the small investors who flocked in mass to the housing market during the period of QE policies conducted by the Federal Reserve Bank.

The top 1% of investors consists of the so-called "Wall Street landlords", that is, private equity-backed investors (i.e., Blackstone Invitation Homes and American Homes 4 Rent). The top 1% also includes the Apartment REITs (i.e., Equity Residential and AvalonBay Communities) that are part of the Real Estate Sector of the Standard and Poor 500 index. While these investors hold a significant share value of the stock of residential capital, their purchases are geographically concentrated. According to our calculations, 90% of the purchases by the top Wall Street Landlords and the public apartment REITs in 2015 are concentrated in 37 MSAs, which is 10% of all MSAs. This investment strategy differs from the small investors that hold a large share in their respective location, and this pattern is observed across all MSAs.

[7] We convert all prices to 2006 dollars using the monthly CPI index.

What is the connection between investment activity and housing affordability? For the period 2009-17, Figure 5 shows a strong correlation between an increase in investors' activity and a worsening in housing affordability. This is essentially stating that in areas that prices increased more relative to income, investors were also more active purchasing housing. Figure A1 in the Online Appendix shows the same correlation in a scatter plot highlighting the population of each MSA.

3 Real Estate Investors and Affordability in the Cross-Section

3.1 Basic Specification

The cross-sectional data showed that the MSAs that experienced the largest increase in the price-to-income ratio post-crisis also had the largest market share of housing purchases by investors. The objective of this section is to study the effect of real estate investors on housing affordability exploiting the cross-sectional differences. The key regression is defined by

$$y_{m,09-17} = \beta_0 + \beta_1 Inv_{m,09-17} + \gamma C_m + \alpha_s + u_m, \tag{1}$$

where $y_{m,09-17}$ denotes the relevant housing variables for a given MSA indexed by m and for the period 2009-17. The relevant housing variables include the average annual real housing price growth rate and the price-to-income ratio for different price-to-income percentiles. To study the effects of investors on the supply of residential units we use the change in construction permits for different types of housing units (i.e., single-family, 2-units, 3-4 units, multiple units). $Inv_{m,09-17}$ is the average share of the investors' dollar value of purchases over the total purchases in MSA m over the same period. The term C_m summarizes traditional MSA-specific controls: population growth, income growth, changes in the unemployment rate, whether the location is sensitive to large house price movements measured by the average real housing price growth during the 2000-2006 boom and the 2006-2007 bust. We also include as a control for the number of building permits in 2007, to account for new supply. The term α_s includes state dummies to account for the time-invariant state-specific influences.

One of the challenges of a direct estimation of specification (1) using OLS is that the parameter of interest could be biased downwards. That would downplay the role of real estate investors capturing "reverse causality" if the investors target MSAs where prices declined the

most after the Great Recession and were slow to pick up. To overcome this potential problem, we use an instrument for the investors' market share of purchases.

3.2 The Instrumental Variable: Propensity to Invest

We use an instrumental variable that allows us to exploit variation in the geographical presence of investors and that is plausibly exogenous to the drivers of housing markets that we cannot control for through fixed effects and local economic variables. As we described in Section 2, this instrument is the average propensity to invest in a given MSA as proxied by the average share of business income by the top earners in an MSA for the year 2007. Top earners are residents that file total income larger than $100,000 in their tax returns. Using the year 2007 assesses the likelihood to make investments in a period with relatively high returns on the risk free rate. With the decline in this rate during the financial crisis, these are the individuals that are most likely to invest in housing to generate a regular cash-flow. In other words, the instrument measures knowledgeable investors with high earnings, prone to invest in real estate. Before the enactment of QE policies at the national level, the areas with a higher share of business income earners were more prone to the search for yield in real estate markers. Consistent with this theory, De Stefani (2021) documents that the investment attitude towards housing increased significantly among the wealthy U.S. population following the financial crisis. There has been a similar increase of investors' activity during the pandemic.

A related channel has been explored for financial investments (i.e., Martínez-Miera and Repullo 2017; Rodnyansky and Darmouni 2017; Campbell and Sigalov 2022; Daniel, Garlappi and Xiao 2021).

Crucially for the validity of our identification strategy, conditional on multiple controls, the geographical distribution of these top earners claiming business income in 2007 is uncorrelated with other factors that drove the appreciation of house prices and price-to-income ratio during the period 2009-17. In other words, it is unlikely that these business individuals picked geographical locations anticipating a decline in house values in 2008 and a future appreciation between 2009-15.

Section 5 contains multiple tests that suggest that the instrument is uncorrelated with other possible factors driving housing markets. One reason is that the baseline specification controls for the key variables that the literature discusses as key drivers of house prices. Another reason is that business income is closely linked to entrepreneurship decisions, and the existing literature finds it extremely challenging to explain geographical differences in entrepreneurship,

which seem related to random historical events (Davidsson 1991; Rocha and Sternberg 2005; Bosma and Kelley 2019). Thus, the evidence suggests that the instrument satisfies the exclusion restriction conditional on the multiple controls.

Table 2 assesses the relevance of the instrument, showing the results of the first stage of the 2-stage least squares (2SLS) regression based on (1). After controlling for the relevant MSA-level controls and state dummies, the instrument is significantly correlated with the investors' purchases. The Wald F statistic of 19.4, reported in 3, allows us to reject that the instrument is weak.

3.3 Results in the Cross-Section

The effects of real estate investors' purchases over the period 2009-17 on price growth, price-to-income, and across price and income tiers are detailed in Table 3. The first column reports the OLS estimation of (1) for the median house price and median income. The smaller coefficient of the OLS estimation is consistent with the expected downward bias of the OLS, since the prices were falling significantly up to 2012, and investors were likely to select areas where prices collapsed. This could be particularly important for large institutional investors that could select the Sun Belt areas where prices suffered one of the largest declines in the nation.

The IV estimation in Table 3 shows that a one-percentage-point increase in the share of investors' purchases leads to a 0.24 percentage points increase in the mid-tier real house price growth. This results is precisely what Lambie-Hanson, Li and Slonkosky (2022) find using a different instrumental variable and geographic unit. Moreover, we show that a one-percentage-point increase in the share of investors' purchases leads to a 0.30 percentage points increase in the bottom-tier real house price growth and a 0.18 percentage points increase in the top-tier real house price growth.

Looking at the standardized estimates, an increase of one standard deviation in real estate purchases by investors (7.78% from Table 1, Panel A) causes 0.83 standard deviations, or 1.46 percentage points, higher housing price growth for the median house.[8] However, the largest effects are estimated for the housing units transacted from the bottom price-tier. In this market segment, an increase in purchases of one standard deviation causes 0.91 standard deviations, or 2.29 percentage points, higher housing price growth.[9]

[8]The standardized estimates use the standardized share of investors and standardized dependent variables, for easier comparison and derivation of the economic significance of the results.We restrict the sample of the standardized variables to the MSAs for which we have Zillow housing prices for all price tiers, to facilitate comparison.

[9]The impact of purchases in the mid-tier market are calculated using 0.827 from 3 multiplied by 1.77 from

What are the effects on affordability? The results for the price-to-income ratio in Table 3 show positive effects from real estate investors. Clearly, the investors had the largest effect in the bottom tier, but also drove prices in the top tier. For example, from Table 1 we know that the average price-to-income ratio in the bottom tier is about 8.5. The estimates indicate that an increase in the level of purchases of one standard deviation would make the price-income ratio increase to a value near 20. Clearly, the affordability impact of the purchase is evident in the bottom price tier.

Recall from the summary statistics (Table 1, Panel A) the average growth in real housing prices between 2009 and 2017 was 0.47%. Our results show that the effect of investors was to prevent even larger drops in housing prices and eventually recover the positive growth.

3.3.1 Does Size and Location Matter?

How much of the results depend on the size and the geographical location of the investors and the investments? In the results in Table 3, both small-local investors and the large Wall Street landlords are active at the same time in some specific MSAs. To isolate the role of small, local investors, we remove from the analysis the specific MSAs in which the Wall Street investors dominate the share of purchases. This ensures that the effects we find are driven by the small, local investors. As we noted previously, the large Wall Street landlords purchase real estate mainly in 10% of the MSAs. Table 4 replicates the analysis from Table 3 for two different sub-samples. The first one excludes the superstar cities, or top 20 MSAs based on the largest 1% investors' purchases. The second excluded the top 37 MSAs where 90% of the value of purchases of the largest 1% investors is geographically located.

The results from Table 4 show that the estimated effect of real estate investors' purchases remains very significant even as we remove the top MSAs. The magnitude in the bottom-tier prices becomes even larger than in the full sample. Quantitatively, one percentage point increase in the share of investors' purchases increases bottom-tier price growth by 0.302 in the full sample, 0.313 in the sample without the top 20 superstar cities, and 0.331 in the sample without the top 37 superstar cities.

The baseline findings are robust to excluding superstar cities, but more importantly show strong evidence on the positive impact of small, local investors on house price growth and the negative impact on affordability.[10] To check the robustness of the results to the geographical

1. Similarly, for the button price-tier the value 0.909 also comes from 3 and it is multiplied by 2.52 from 1.

[10]To classify the superstar cities, we rank the MSAs based on the dollar value of purchases by the largest 1% investors. The top MSAs are the superstars. We perform two robustness checks, ranking the MSAs by (a)

unit, we perform the same analysis with counties instead of MSAs. Table $A1$ shows that the results remain unchanged when we use counties.

3.3.2 Real Estate Investors and New Construction: Single vs. Multi-Family

The stock of residential housing evolves very slowly over time. The type of units that are being demanded by the market are the units that developers will try to supply with a lag. During the boom in 2003-06, a large share of the demand came from individual homeowners that wanted to purchase single-family housing. The tightening of credit standards after the financial crisis coincided with a period of adjustment of the households' balance sheets (i.e., see Garriga and Hedlund 2020 for a detailed quantitative analysis using a model of household purchases and endogenous house prices). The declining demand of owner-occupied housing changed the type of newly constructed units. Table 5 summarizes the impact of real estate investors on new construction over the period 2009-17. The first column reports the IV estimation of (1) for the total number of construction permits for all houses, measured in logs. According to the estimates, a one percentage point increase in the share of investors increases the number of new construction permits by 5.2% on average ($e^{0.051} - 1$).

How do these purchases impact the characteristics of the stock of residential housing? The analysis that separates permits for single-family and multi-family units indicates that investor purchases lead to an increase in permits of 4.8% ($e^{0.047} - 1$) for single-unit houses, and an average increase of 16.4% ($e^{0.152} - 1$) for buildings of 5 or more units. Table $A2$ in the Appendix reiterates the results when the superstar cities are eliminated from the sample. This indicates that the composition of the demand has lasting effects in the type of newly constructed residential structures put in place. Depending on the degree of persistence of the shock that changes the composition of the housing demand (i.e., a transitory or a permanent change in the fraction of households desiring to enter in the owner-occupied market), the stock of housing might evolve in one direction or another. Since the characteristics of the stock of housing change very slowly, the type of unity newly constructed between 2009-17, mainly multi-family and top-tier, can rationalize the lack of availability of single-family houses during the pandemic.

the share of purchases by the largest 1% investors over the total purchases by investors and (b) the share of purchases by the largest 1% investors over the total purchases by investors and households. The alternative classifications have large overlaps with the first definition of superstar cities and they don't change the results.

4 Dynamic Real Effects of Investors

This section expands the analysis in Section 3 by studying the dynamic effects of the investors' purchases over time and across the geography. We follow Jordà (2005) and estimate sequential regressions of the dependent variable shifted forward.[11] The dynamic specification is defined by

$$y_{m,t+i} = \beta_0 + \beta_1^{(i)} Inv_{m,t-1} + \beta_2 y_{m,t-1} + \gamma C_{m,t-1} + \alpha_m + b_t + u_{m,t}, \tag{2}$$

where t indexes years and m MSAs, and $y_{m,t}$ denotes the housing variables: real housing price growth rate from year $t-1$ to year t, for all price tiers, the price-to-income and rent-to-income ratios, the price-to-rent ratio, and new construction permits. $Inv_{m,t-1}$ is the investors' share of dollar value of purchases over the total market value for the year $t-1$ in MSA m. The term $C_{m,t-1}$ captures time-varying MSA-specific controls (the population growth rate, the median income growth rate, and the change in the unemployment rate).[12] The location fixed effects α_m capture the time-invariant MSA-specific influences, and the time fixed effects b_t account for the time-varying factors common to all MSAs, like national mortgage rates. We include a lagged dependent variable $y_{m,t-1}$ to allow the growth response to be temporary.

The estimate of interest is the vector of $\{\beta_1^{(i)}\}$, where $i = 0, 1, ..., 6$ is the time horizon of the response, that is, the number of years after the investors' purchases. Each $\beta_1^{(i)}$ corresponds to the effect of investors' share of purchases at horizon i. Setting $i = 0$ gives the usual panel specification. We estimate (2) for the full panel data from 2009 to 2017. In the estimation we cluster standard errors by MSA to allow for within-MSA correlation throughout the sample period.[13]

The dynamic nature of the analysis requires adjusting the instrument. The notion of the propensity to invest is determined using the cross-sectional information in 2007 before the financial crisis and before the recovery of housing markets. For specification 2, we interact the previous instrument with the time path of certificate of deposits (CD) interest rates. The key idea is to exploit the unexpected nature of QE that triggered a national shock to the CD rate, which is equal for all locations and it is not driven by local factors. The exposure of each location to the national shock is unrelated to local factors affecting the housing markets, as we assess in Section 5. The exposure is also predetermined, fixed in 2007, which minimizes the possibility of

[11] Favara and Imbs (2015) also apply this method to study house prices, and Mian, Sufi and Verner (2017) to study GDP growth.

[12] Controlling for contemporaneous income and population growth, and unemployment rate change doesn't change the results.

[13] The results remain unchanged when we alternatively allow for Newey-West standard errors that allow for heteroskedasticity and within-MSA serial autocorrelation of the error term.

reverse causality. Thus, this instrument captures which MSAs are more likely to have housing investors after the QE policies. The rationale is analogous to the housing net worth channel of Mian and Sufi (2014) that exposes certain areas to larger macro effects from declines in housing prices due to their housing leverage. In our case, we expose investment-prone areas to the QE shock. Table $A3$ shows that the relevance condition is satisfied.

Figures 6 to 8 display the baseline findings.[14] On impact, the purchases of real estate investors have a positive impact on price and rent growth. Over time, the effect on growth de-accelerates and eventually. For house prices this is around 3 whereas for rents the momentum stops in year 4. The increase in prices makes the cap rates decline rationalizing the flattening in the growth of investor's purchases. However, the cumulative effects on prices and rents are positive and large as can be seen in Figure 6. What is the impact on affordability? Figure 7 shows an average result across MSAs for the effects of investors' purchases on price-to-income and rent-to-income ratios. Clearly, most of the effects on affordability happened in the initial 3 to 4 years. Once the employment started to recover the measures of affordability improved from the perspective of the impact of the investors.

How do purchases of real estate investors affect the timing and the type characteristics of newly constructed units? The larger short-run effect in the panel regressions, relative to the cross-section findings, can be rationalized by the timing of the response of new construction. As can be seen in Figure 7 bottom left panel, new construction measured by building permits has a hump shape response with a peak around 2-3 years that then remains elevated for several years. The purchases of investors and the implied price growth motivates a supply response that partially mitigates the negative impact on affordability. Permits measure expectations about future growth, and construction developers respond to that incentive. The very short-run response in the supply can be assessed by analyzing the evolution of vacancies. The bottom right panel of Figure 7 highlights that as investors are attracted to the currently available housing units for sale the number of vacancies declines. Over time, as the cost of residential units increases, cap rates decrease, and vacancies increase as newly constructed units arrive to the market.[15]

Are there key differences in the cross-section once we consider the difficulties in certain areas to rapidly expand the supply? There are some striking differences in the response of price-to-income and price-to-rent once we split the sample of MSAs by the housing supply elasticity as can be seen in Figure 8. In highly inelastic areas, the short-run price response and the implied

[14] Tables $A4$ and $A5$ have the results of the estimation.

[15] Ben-David, Towbin and Weber (2019) argue that one way to identify housing booms is to look at the response of vacancies for owner-occupied and rental houses.

worsening of affordability are much larger than in MSAs with high supply elasticity. In other words, in these areas with low supply elasticity, investors drive prices and don't seem to move rents in the short-run. As a result the price-to-rent ratio increases, the price-to-income ratio also increases, and the rent-to-income ratio is constant. In areas with high supply elasticity the opposite effect is true. The price-to-rent ratio decreases in the short run and most of the effect on affordability comes from rents and not prices. The distributional effects are very different from the average effects depicted in Figure 7. There is a clear separation in the response of prices and rents across elasticities.

The final part of the analysis explores the impact across housing market segments and housing characteristics. Consistent with the cross-sectional evidence, the analysis indicates that investors have larger effects on the bottom-tier of the market. The differences in the estimated impulse responses for the top and bottom price tier as depicted in Figure $A2$. While multi-family can be important for the supply of new units, it is important to highlight the role of the single-family segment of the housing market. We redo the analysis separating these units from the rest of the market. The top panel of Table $A6$ uses single-family prices and the bottom panel prices for all homes, from Zillow. The findings indicate that the response of prices to investors is exactly as statistically significant in the single-family segment as in the total market.[16]

5 Validity of the Instrument

In this section we assess the validity of the instrumental variable and the robustness of the previous results. We examine at length the exclusion restriction. Section 3.2 already discussed the relevance of the instrument. Figures $A3$ and $A4$ confirm that the instrument is strongly correlated with the investors' share of purchases.

Our instrumental variable measures the exposure of each MSA to the national sudden drop in interest rates. The identification concern is whether differences across MSAs in the share of income reported as business income in 2007 by high-earners leads to differential changes in the outcome variables through channels other than investors and for which our fixed effects and control variables cannot control for.

We follow different strategies to test the exclusion restriction: 1) Our empirical design satisfies the parallel pre-trends. Placebo tests confirm the parallel pre-trends. This is strong

[16] Ninety percent of the properties in the Zillow Home Value Index are single-family and the rest are condominiums and cooperatives.

support for the plausibility of the exogeneity assumption according to Goldsmith-Pinkham, Sorkin and Swift (2020). 2) We control exhaustively for all the usual drivers of housing markets (income, local economy activity, credit conditions, population, composition of labor markets etc.) through different variables. Even if they can be "bad control variables" that should not be in the main specification because they are part of the transmission channel of investors' purchases. None changes the main results. Thus, it does not seem that the usual drivers of housing markets are driving our main results as omitted variables. We were not able to think on extra omitted variables driving both housing markets and the cross-sectional differences in our instrument. 3) Altonji, Elder and Taber (2005) and Oster (2019) omitted variable tests suggest that there is no omitted variable bias. 4) We show that it is very hard to predict the instrument, which indicates that a large part of the variation in it is random, especially unrelated to other drivers of housing markets. 5) We show the robustness of the results to alternative specifications and definitions of the investors' purchases.

5.1 Parallel pre-trends

The use of a Bartik-like instrument and the availability of pre-period trends, make our empirical strategy analogous to difference-in-differences. In a difference-in-differences setting the MSAs with the largest exposure to business income of top earners in 2007 is the treated group, and the MSAs with the smallest exposure is the control group. The year 2008 is the "treatment" year, when the Fed implemented the first wave of unconventional monetary policy which led to a large drop in interest rates.

Figure 9 plots the annual log number of building permits and the annual real price growth of bottom-tier homes for MSAs ranking in the top and bottom 25% of exposure to top earners' business income in 2007. Figure 9 shows that, prior to the shock, the high and low exposure groups have parallel dynamics. The divergence starts post-2008. That is, the MSAs behave similarly in the period when QE does not exist. We only see differences during and after the QE period when the MSAs more exposed to potential investors see those investors move to the housing market in search for yields. Thus, the parallel pre-trends suggest that the instrument is driving construction and prices only in the post-crisis period. In other words, the instrument is not capturing other factors that could make housing prices to have permanently different dynamics across locations. Goldsmith-Pinkham, Sorkin and Swift (2020) recommend this test to assess whether the exclusion restriction is valid.

5.2 Placebo tests

Another way to implement the parallel trends test recommended by Goldsmith-Pinkham, Sorkin and Swift (2020) is to do a placebo analysis. Figure $A5$ does such a placebo test with the pre-crisis period 2000-2006 when QE was not operating. The scatterplots control for the same variables as specification (1). The MSAs are binned by percentiles so that each point represents around 15 MSAs. The bottom panel of the figure demonstrates strong positive correlation between the instrument and housing price growth over 2009–2017. This correlation is absent in the pre-crisis placebo sample that is in the top panel. This evidence suggests that the instrument is not contaminated by pre-crisis price growth.

To confirm the message from Figure $A5$, we conduct various placebo tests over the 2000–2006, 2001–2006, and 2000–2005 periods in Table 6.[17] We ask if, when using a specification analogous to (1), the exposure to the top earners' business income can explain housing price growth over any of these periods. The placebo point estimates are insignificant across periods. That is, the instrument is only capturing post-crisis positive shocks in housing investment. None of the factors operating pre-QE period are correlated with the instrument.

Table $A7$ contains the results of placebo tests for the panel analysis, for pre-crisis periods. Figure $A6$ plots a placebo experiment linking the instrument to prices, and Figure $A7$ to new construction. The instrument does not contribute to changes in prices or number of construction permits in time periods pre-crisis.

5.3 Controls for the local economy

To rule out the possibility that local economic conditions drive the results, Table 7 re-estimates the baseline specification controlling for a wide range of variables that capture contemporaneous local economic activity: average annual unemployment rate change, labor force participation growth, real GDP per capita growth, and median hourly wage per capita growth from 2009 to 2017. Table 7 displays results very similar to Table 3. Importantly, the estimated coefficients are in a close range of the baseline coefficient of 0.234 from Table 3. A large change in the coefficient would hint at omitted variables biasing the estimation. These results suggest that the local economic activity and the investors are both important for housing price growth, but investors also affect housing markets even when keeping local economic activity constant.

[17] The selection of placebo periods is restricted by a lower bound of the year 2000, since this is when our investors' data begin. The upper bound is 2006, since we want to avoid an overlap and potential co-determination of the investors' share and our instrumental variable that is constructed using 2007 data.

5.4 Controls for credit conditions

Credit conditions are another potential driver of housing prices that we want to rule out. Table 8 reestimates the baseline specification including controls for credit supply. The first column controls for the mortgage denial rate over 2009-2017 in each MSA. The second column controls for the share of lenders, in terms of their deposit holdings, that underwent stress-testing due to the Dodd-Frank Act. This control is inspired by Gete and Reher (2018) who use this variable as an instrument for denial rates to study housing rents. After including those controls, the coefficient of the investors' share moves between 2% and 8%, and remains strongly significant.

5.5 Controls for shifts in the composition of labor demand

Although we include several controls for economic conditions, an alternative concern could be that the instrument is correlated with the industrial composition of the local labor market, and therefore related to shifts in the composition of labor demand during the post-crisis period.[18] To address this concern we reestimate the baseline specification controlling for changes in employment in the largest industry sectors within the MSAs (Table 9). The changes are accounted for, starting from the base year of the instrumental variable, that is, from the annual change from 2007 to 2008, up to the annual change from 2016 to 2017. Employment changes in some industries, such as Real Estate, Rental and Leasing could be considered bad controls, as they are likely part of the transmission channel of investors on prices. Even with this prudent analysis, after controlling for employment growth of up to ten industries, the estimated effect of investors holds, and it is close to the baseline effect.

Moreover, Table *A*8 re-estimates the dynamic results accounting for the lagged annual shifts in the composition of labor demand. The dynamic patterns of housing price growth remain unchanged when we include the employment growth controls for the largest industries in the MSAs. The shifts in the composition of labor demand do not seem to be driving the results.

5.6 Oster (2019) omitted variable bias test

We conduct omitted variable bias tests based on the work of Altonji, Elder and Taber (2005) and Oster (2019), which we outline in the online Appendix *C*. This test confirms that, while our

[18] For example, Monte, Redding and Rossi-Hansberg (2018) study the importance of spatial spillovers due to local labor demand shocks through changes in commuting patterns.

multiple controls do not change our coefficient of interest in a significant way, they do increase significantly the R-squared of the estimation.

Table *A*9 shows the results of the omitted variable bias test for four different specifications: (1) our baseline specification in Table 3, (2) the specification with additional controls for economic drivers in Table 7, (3) the specification with additional controls for credit conditions in Table 8 and (4) the specification with additional controls for changes in industry employment in Table 9. The results strongly reject that the effect of the share of investors on housing prices is driven by omitted variables. Thus, these tests alleviate concerns of omitted variable bias.

5.7 Unpredictable instrumental variable

Here, we show that it is very difficult to predict the share of business income. In the introduction we discuss papers showing that most of the cross-regional differences in investment attitude are as good as random. It is very hard to predict the investment or entrepreneurship attitude of an MSA. We confirm this result in Table 10. We regress the share of the top earners' business income in each MSA in 2007 on several factors that may explain investment or entrepreneurship activity. These factors are demographic (median age and share of immigrants), regulatory (tax rate for high earners), geographical (natural amenity index) and the ranking of MSAs in the ease of doing business. While some of these factors are correlated with the top earners' business income, their explanatory power is low. The demographic and regulatory factors explain 11% of the variation in the top earners' business income share, as we see by the R-squared of the first column of Table 10. Including the geographical factor the R-squared becomes 22%.

Moreover, in Table *A*10 we study whether the standard drivers of the housing market are correlated with the instrument, given our controls. We regress the local share of top earners' business income on the pre-crisis trends of homeownership and median age within each MSA. To better gauge the magnitude of these partial correlations, the table normalizes all variables to have a mean of zero and a variance of one. This allows us to assess both the magnitude and statistical significance of any correlations. Importantly, there is no relevant correlation between the common drivers of housing variables and the MSA share of top earners' business income.

5.8 Robustness to other specifications

We check robustness to changes in the specifications. First, we use additional controls for total demand for housing or demand for housing by investors. These controls are the total dollar

value of purchases in the market or the total dollar value of purchases by investors. Controlling for either of these levels of demand does not change any of the results.[19] Our baseline controls (population, income, unemployment, MSA and year fixed effects) already capture a large part of the variation in housing demand. Second, we use an additional control for the share of purchases by individual investors in the housing market of each MSA. We identify individual investors as individuals (having their personal name in the deeds) who purchase two or more houses in the same MSA within two years. Table $A11$ shows that the main effects we study remain unchanged after the inclusion of this control. The share of purchases by individual investors shows correlation with price growth, and this is in addition to the effects of the investors who purchase houses as legal entities. Third, we control for the change in the share of foreclosures in each MSA. Foreclosed properties are likely to attract investors because of lower prices, and at the same time they might restrict access to investors in some areas through the Fannie Mae and Freddie Mac First Look programs (Lambie-Hanson, Li and Slonkosky 2022). This analysis uses a restricted sample of 84 MSAs for which we have foreclosure data from Zillow for the years 2008 to 2017. Even with this restricted sample, the effects of investors on house price growth remain significant for all price tiers, as Table $A12$ shows. Finally, Table $A13$ shows that our results are robust to using alternative measures of investors' share based on number of purchases and number of units.

6 Conclusions

In this paper we analyze the contribution of real estate investors to the U.S. residential housing markets post Great Recession. Using a large database covering the whole U.S. we document the emergence of a new type of real estate investors that buys properties through legal entities. These investors are local, relatively small in size, and present in MSAs all across the U.S. Instead, large investors, as those referred to as Wall Street Landlords, are geographically concentrated in "superstar cities". The growth of small and local investors in both extensive (number of investors) and intensive (dollar purchases) margins in the post-financial crisis period is astonishing.

Then, we analyze how the real estate investors affect housing affordability. Cities around the world are designing policies to deal with these new investors. Investors drove most of the recovery in housing prices, especially in low-tier housing, and housing affordability worsened in the short term. Especially affected were the single-family homes at the bottom of the price distribution. These are usually starter homes that otherwise would be purchased by young

[19] We do not report the tables of these results, as they are similar to the previous ones. Available upon request.

households. The presence of investors triggered an equilibrium response of supply, which lead to improving affordability in the medium term. The investors affected differently the price and rent affordability, depending on the supply restrictions of each area. Price increases were small in MSAs where there are loose supply restrictions.

References

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S. and Evanoff, D.: 2019, Mitigating investor losses due to mortgage defaults: Lessons from the Global Financial Crisis. In Franklin Allen, Ester Faia, Michalis Haliassos, Katja Langenbucher, eds., Capital Markets Union and Beyond, MIT Press.

Albanesi, S., De Giorgi, G. and Nosal, J.: 2017, Credit growth and the financial crisis: A new narrative.

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.

Altonji, J. G., Elder, T. E. and Taber, C. R.: 2005, Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools, *Journal of Political Economy* **113**(1), 151–184.

Bayer, P., Mangum, K. and Roberts, J. W.: 2021, Speculative fever: Investor contagion in the housing bubble, *American Economic Review* **111**(2), 609–51.

Ben-David, I.: 2011, Financial constraints and inflated home prices during the real estate boom, *American Economic Journal: Applied Economics* **3**(3), 55–87.

Ben-David, I., Towbin, P. and Weber, S.: 2019, Expectations during the U.S. housing boom: Inferring beliefs from vacant homes.

Ben-Shahar, D., Gabriel, S. and Oliner, S. D.: 2020, New research on housing affordability, *Regional Science and Urban Economics* **80**, 1–4.

Bernstein, A., Gustafson, M. T. and Lewis, R.: 2019, Disaster on the horizon: The price effect of sea level rise, *Journal of Financial Economics* **134**(2), 253–272.

Bloomberg: 2018, The big problem with investing in Amsterdam's hot housing market. https://www.bloomberg.com/news/articles/2018-08-08/the-big-problem-with-investing-in-amsterdam-s-hot-housing-market.

Bloomberg: 2019, Spain is latest battleground for global affordable housing. https://www.bloomberg.com/news/articles/2019-06-19/spain-is-latest-battleground-in-global-affordable-housing-fight.

Bosma, N. and Kelley, D.: 2019, Global Entrepreneurship Monitor 2018/2019 Global Report. Global Entrepreneurship Research Association.

Brunson, S. and Buttimer, R.: 2020, Wall Street, Main Street, your street: How investors impact the single-family housing market.

Business Insider: 2019, California becomes the third state nationwide to pass a rent control bill to address its affordable housing crisis. https://www.businessinsider.com/california-pass-rent-control-bill-to-address-affordable-housing-crisis-2019-9?IR=T.

Campbell, J. Y. and Sigalov, R.: 2022, Portfolio choice with sustainable spending: A model of reaching for yield, *Journal of Financial Economics* **143**(1), 188–206.

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

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.

Davidsson, P.: 1991, Continued entrepreneurship: Ability, need, and opportunity as determinants of small firm growth, *Journal of Business Venturing* **6**(6), 405–429.

De Stefani, A.: 2021, House price history, biased expectations, and credit cycles: The role of housing investors, *Real Estate Economics* **49**(4), 1238–1266.

Favara, G. and Imbs, J.: 2015, Credit supply and the price of housing, *The American Economic Review* **105**(3), 958–992.

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

Ganduri, R., Xiao, S. C. and Xiao, S. W.: 2019, Tracing the source of liquidity for distressed housing markets.

Garriga, C., Gete, P. and Tsouderou, A.: 2021, Search for yield in housing markets.

Garriga, C. and Hedlund, A.: 2020, Mortgage debt, consumption, and illiquid housing markets in the Great Recession, *American Economic Review* **110**(6), 1603–34.

Gete, P. and Reher, M.: 2018, Mortgage supply and housing rents, *The Review of Financial Studies* **31**(12), 4884–4911.

Goldsmith-Pinkham, P., Sorkin, I. and Swift, H.: 2020, Bartik instruments: What, when, why, and how, *American Economic Review* **110**(8), 2586–2624.

Graham, J.: 2020, House prices, investors, and credit in the Great Housing Bust.

Gurun, U. G., Wu, J., Xiao, S. C. and Xiao, S. W.: 2022, Do Wall Street landlords undermine renters' welfare?, *The Review of Financial Studies* .

Gyourko, J., Mayer, C. and Sinai, T.: 2013, Superstar cities, *American Economic Journal: Economic Policy* **5**(4), 167–99.

Jenwittayaroje, N. and Jiraporn, P.: 2019, Do independent directors improve firm value? Evidence from the Great Recession, *International Review of Finance* **19**(1), 207–222.

Jordà, Ò.: 2005, Estimation and inference of impulse responses by local projections, *American Economic Review* **95**(1), 161–182.

Kleibergen, F. and R. Paap: 2006, Generalized reduced rank tests using the singular value decomposition, *Journal of Econometrics* pp. 97–126.

Lambie-Hanson, L., Li, W. and Slonkosky, M.: 2022, Real estate investors and the U.S. housing recovery, *Real Estate Economics* .

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

Mian, A. and Sufi, A.: 2014, What explains the 2007–2009 drop in employment?, *Econometrica* **82**(6), 2197–2223.

Mian, A., Sufi, A. and Verner, E.: 2017, Household debt and business cycles worldwide, *The Quarterly Journal of Economics* **132**(4), 1755–1817.

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.

Molloy, R., Nathanson, C. G. and Paciorek, A.: 2022, Housing supply and affordability: Evidence from rents, housing consumption and household location, *Journal of Urban Economics* **129**, 1–21.

Monte, F., Redding, S. J. and Rossi-Hansberg, E.: 2018, Commuting, migration, and local employment elasticities, *American Economic Review* **108**(12), 3855–90.

Oster, E.: 2019, Unobservable selection and coefficient stability: Theory and evidence, *Journal of Business & Economic Statistics* **37**(2), 187–204.

Rocha, H. O. and Sternberg, R.: 2005, Entrepreneurship: The role of clusters theoretical perspectives and empirical evidence from Germany, *Small Business Economics* **24**(3), 267–292.

Rodnyansky, A. and Darmouni, O. M.: 2017, The effects of quantitative easing on bank lending behavior, *The Review of Financial Studies* **30**(11), 3858–3887.

Saiz, A.: 2010, The geographic determinants of housing supply, *The Quarterly Journal of Economics* **125**(3), 1253–1296.

Smith, P. S. and Liu, C. H.: 2020, Institutional investment, asset illiquidity and post-crash housing market dynamics, *Real Estate Economics* **48**(3), 673–709.

Stroebel, J.: 2016, Asymmetric information about collateral values, *The Journal of Finance* **71**(3), 1071–1112.

The Wall Street Journal: 2019, In Berlin, a radical proposal to combat rising rents: Expropriate big landlords. https://www.wsj.com/articles/in-berlin-a-radical-proposal-to-combat-rising-rents-expropriate-big-landlords-11554202800.

Zillow: 2017, ZTRAX: Zillow transaction and assessor dataset, 2017-Q4. Zillow Group, Inc. http://www.zillow.com/ztrax/.

Figures

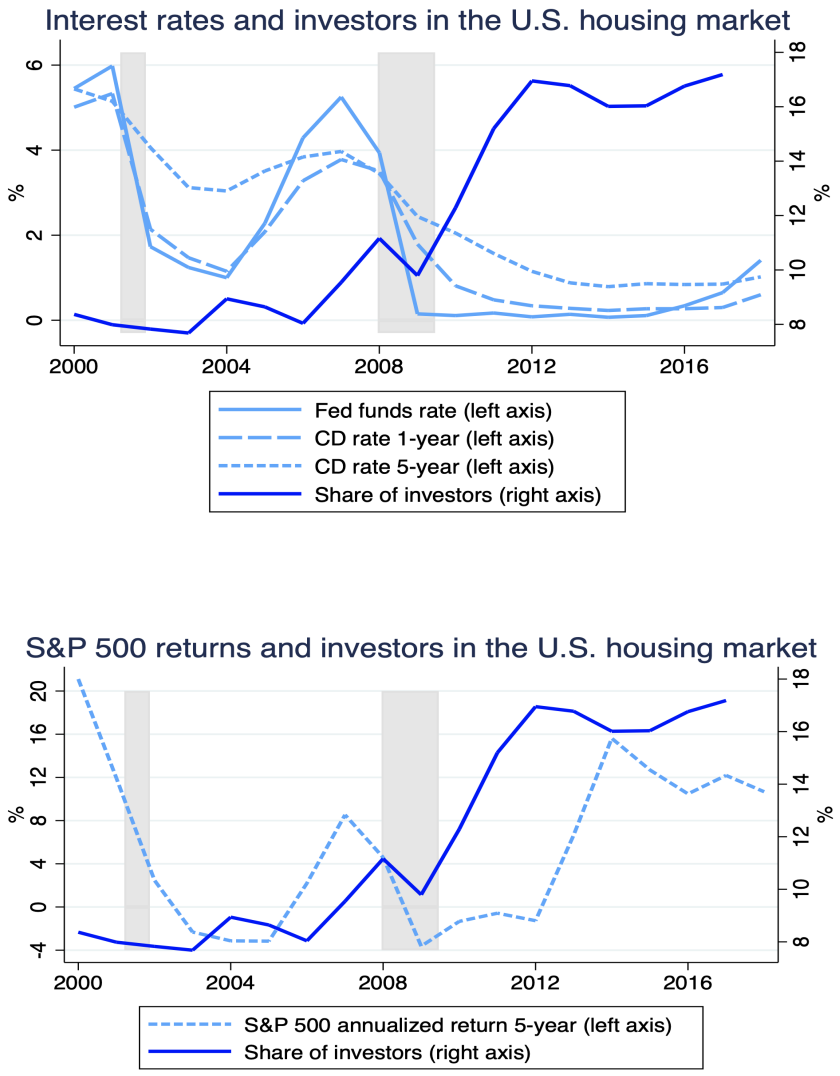

Figure 1. Rates of return and housing investors. The top figure plots the federal funds rate and the average CD rates for 1-year and 5-year CDs. The bottom figure plots the 5-year annualized past returns of the S&P 500 index. Both figures also plot the share of dollar purchases that corresponds to investors in the U.S. housing market. The gray areas illustrate the U.S. Recessions.

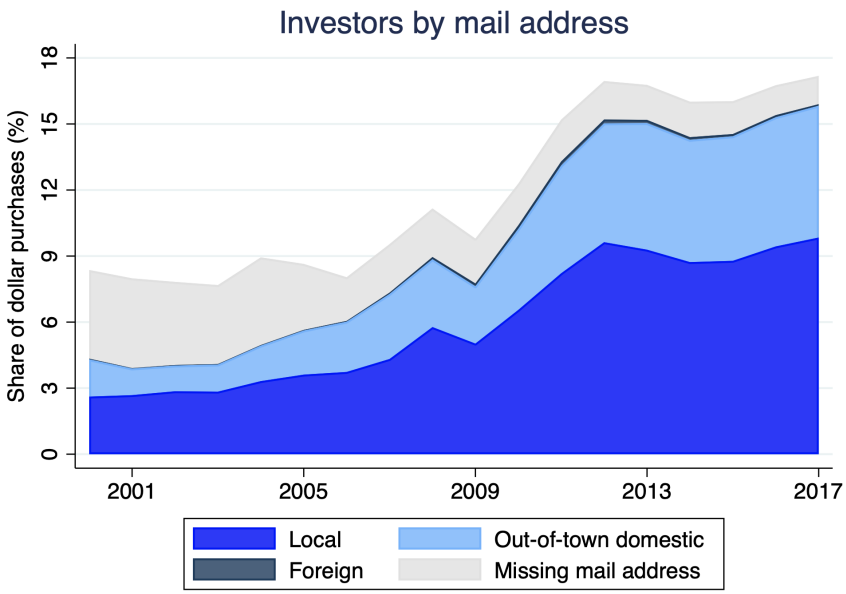

Figure 2. Local and out-of-town investors. The figure plots the share of dollar purchases by investors in the U.S. housing market split into local, out-of-town domestic and foreign investors. Local investors have their mailing address in the same MSA as the property they purchase. Out-of-town domestic investors have their mailing address in the U.S., but outside the MSA of the property they purchase. Foreign investors have a mail address outside the U.S. The property and mailing addresses come from the ZTRAX database.

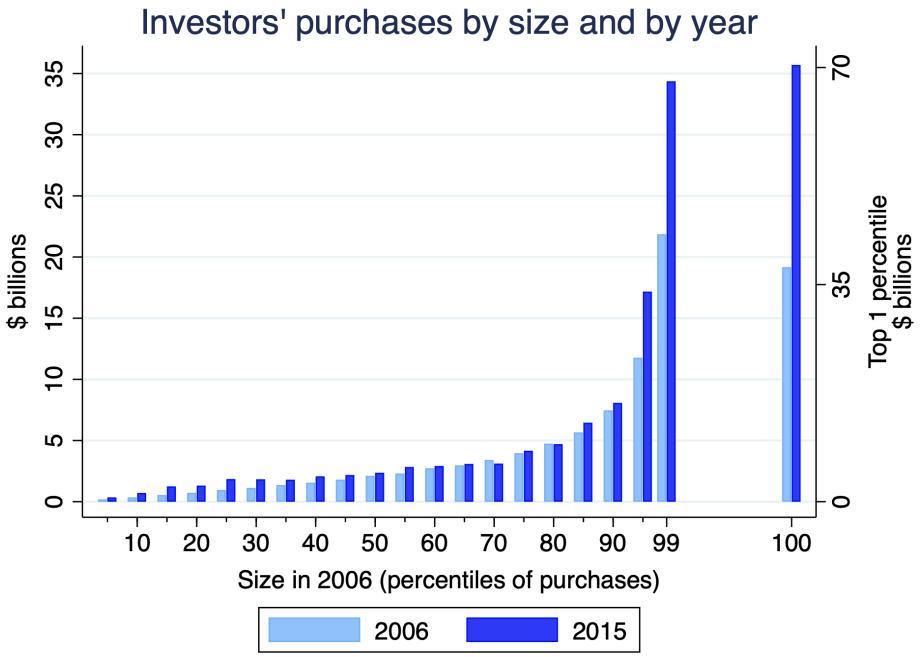

Figure 3. Size distribution of real estate investors. The figure plots the total dollar purchases by real estate investors in the U.S. housing markets in the years 2006 and 2015, in each percentile segment of purchase value. The percentile cutoffs are the dollar values of the cutoffs in 2006. All dollar values are in 2006 dollars.

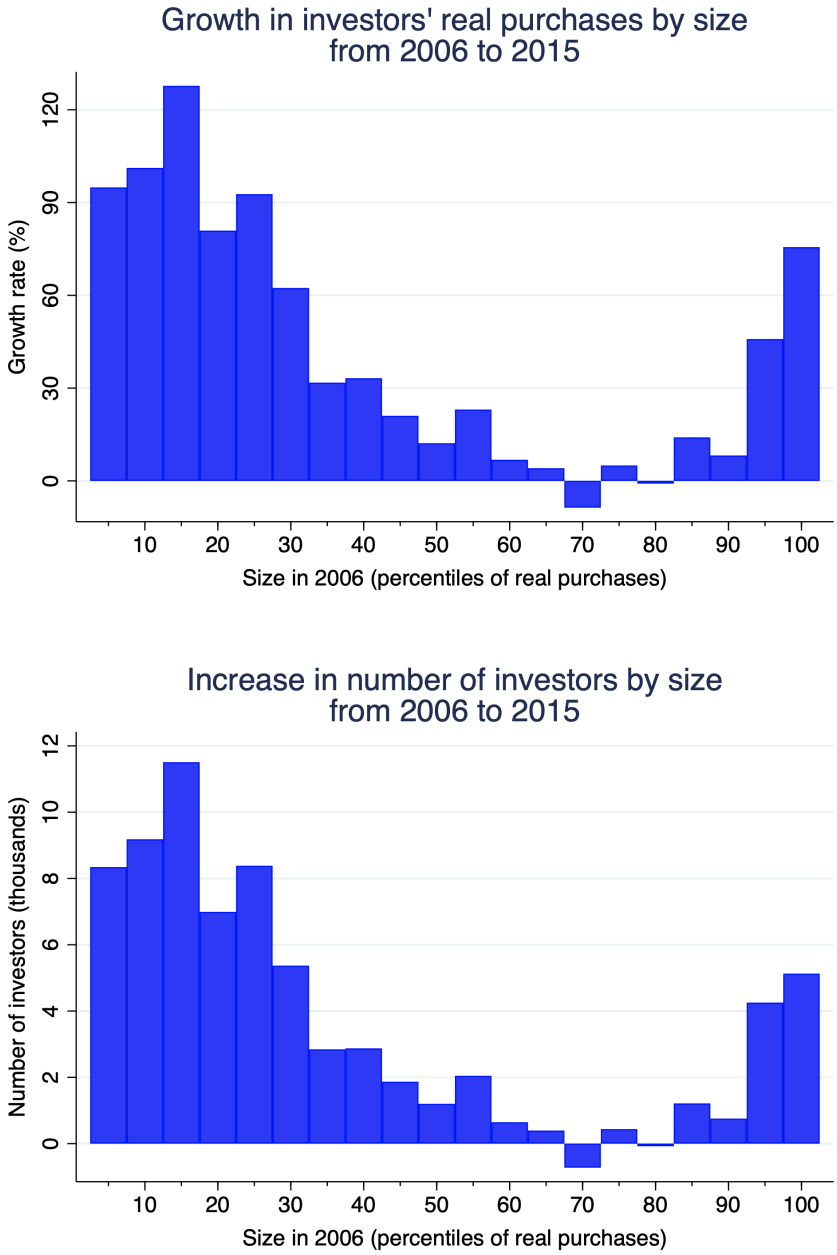

Figure 4. Growth of investors by size. The top figure plots the growth in dollar purchases by investors from 2006 to 2015 in each percentile segment of purchase value. The percentile cutoffs are the dollar values of the cutoffs in 2006. All dollar values are in 2006 dollars. The bottom figure shows the change in the number of investors (extensive margin) over the same period.

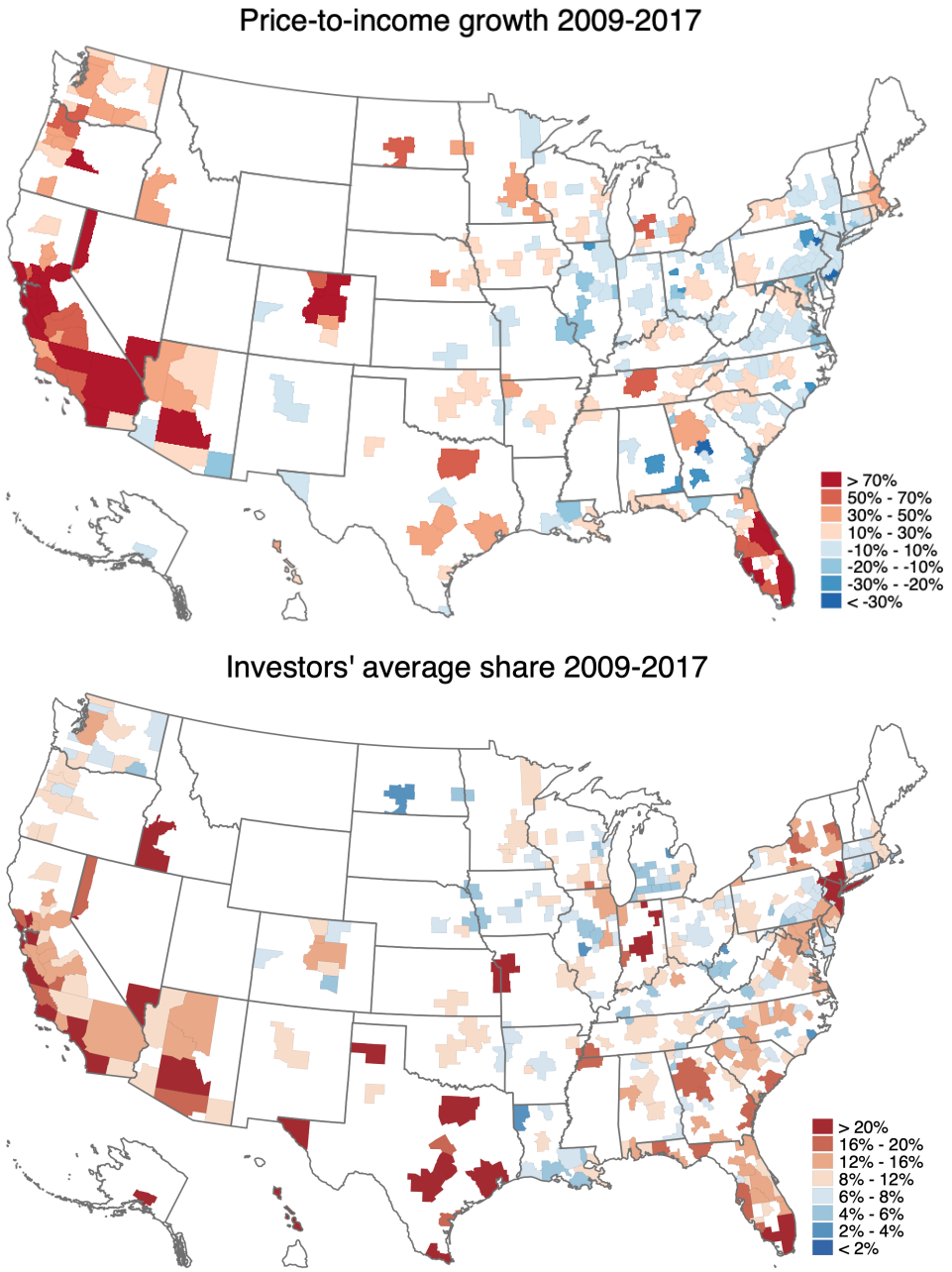

Figure 5. Affordability and investors in the U.S. The top map shows the percentage growth of price-to-income ratio from 2009 to 2017 in each MSA for bottom-tier houses. The bottom map shows the average market share of dollar purchases by investors from 2009 to 2017 in each MSA. Figure A1 shows the correlation of the raw data in a scatter plot.

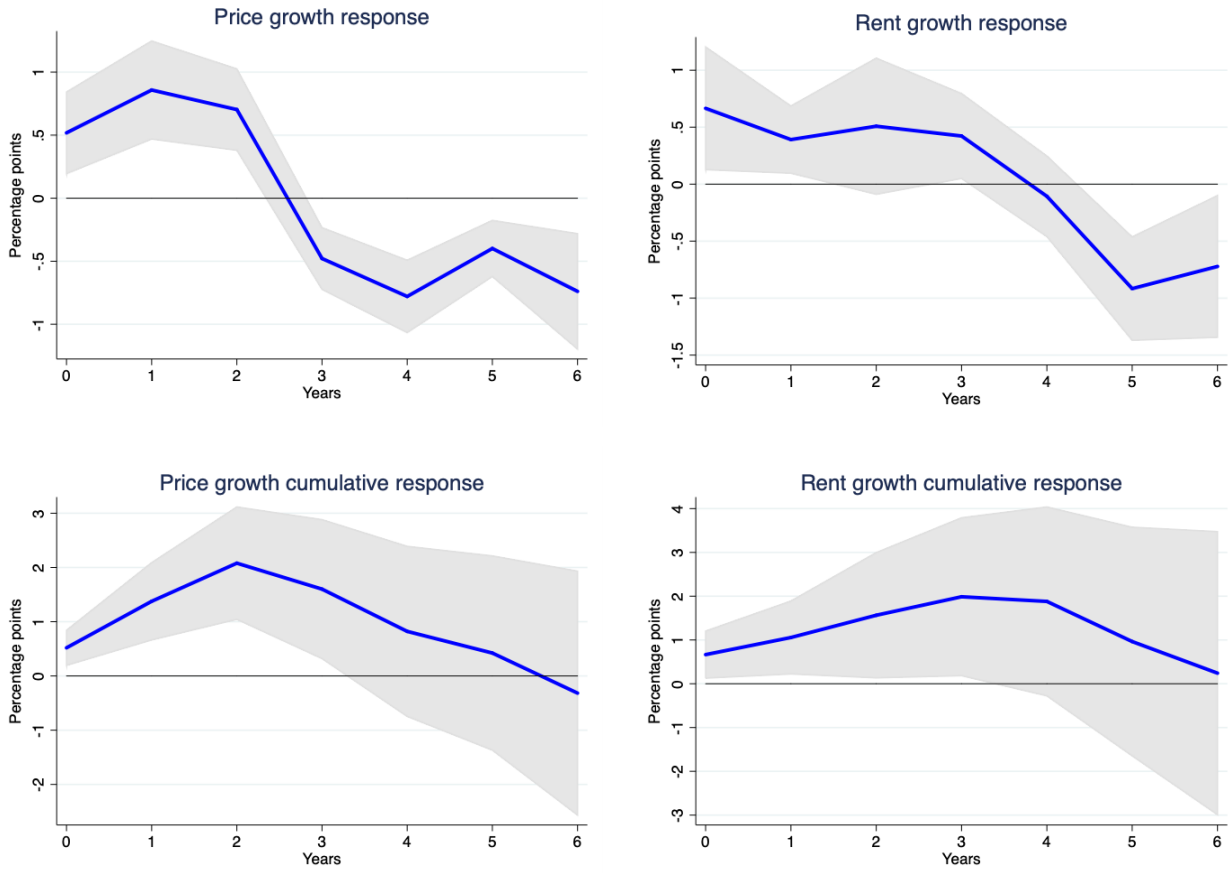

Figure 6. Dynamics of housing prices and rents after investors' purchases. The top figures plot the estimates from sequential regressions of the price growth and rent growth on the instrumented past investors' share of purchases. The bottom figures plot the cumulative effects, calculated as the cumulative sum of the previous coefficients. Prices and rents are adjusted for inflation. Section 4 discusses the methodology that follows Jordà (2005). We estimate the impulse responses for the full panel data from 2009 to 2017. The shaded areas show the 90% confidence interval.

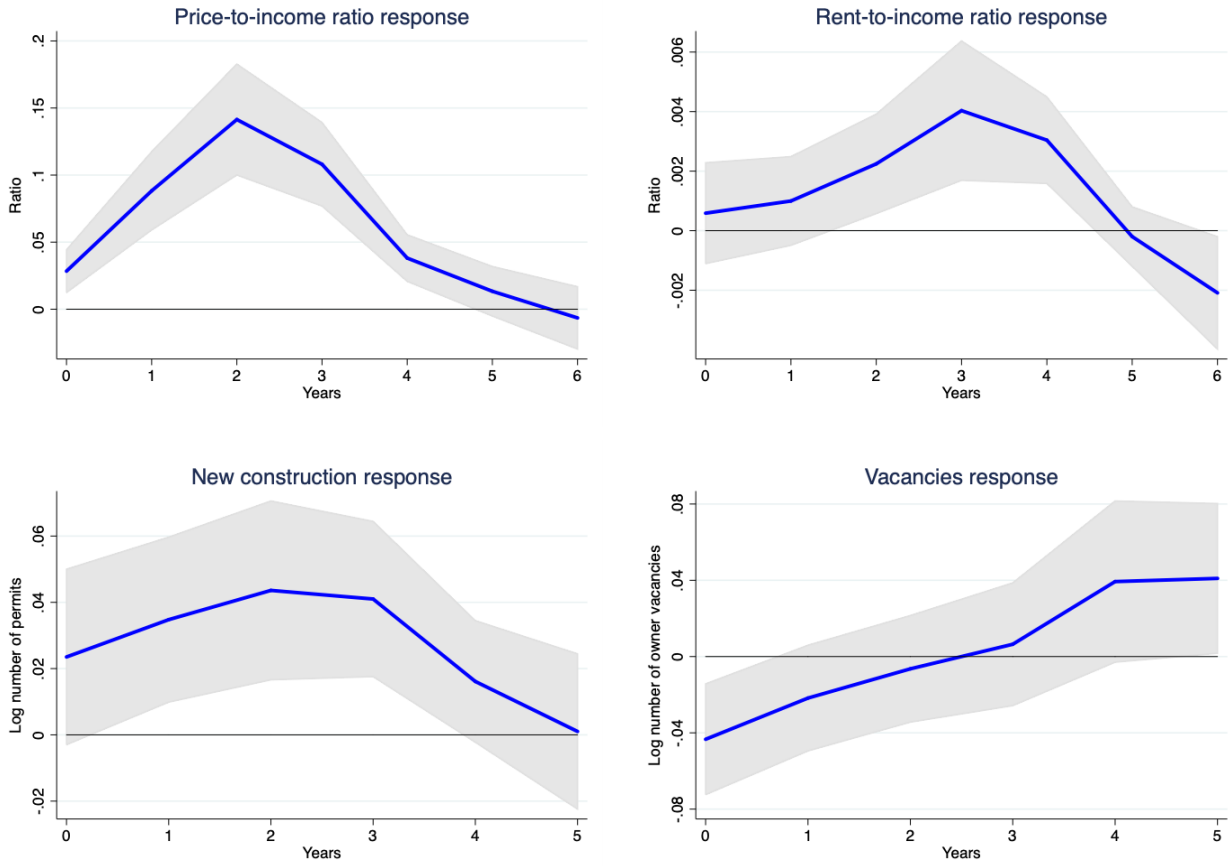

Figure 7. Dynamics of housing affordability and supply after investors' purchases. The figures plot the estimates from sequential regressions of (a) the price-to-income ratio, (b) the rent-to-income ratio, (c) the log number of building permits, and (d) the log number of homeowner vacant units on the instrumented past investors' share of purchases. The price-to-income ratio is the median housing price over the median annual household income in an MSA. The rent-to-income ratio is the median annual housing rent over the median annual household income in an MSA. We estimate the impulse responses for the full panel data from 2009 to 2017. The shaded areas show the 90% confidence interval.

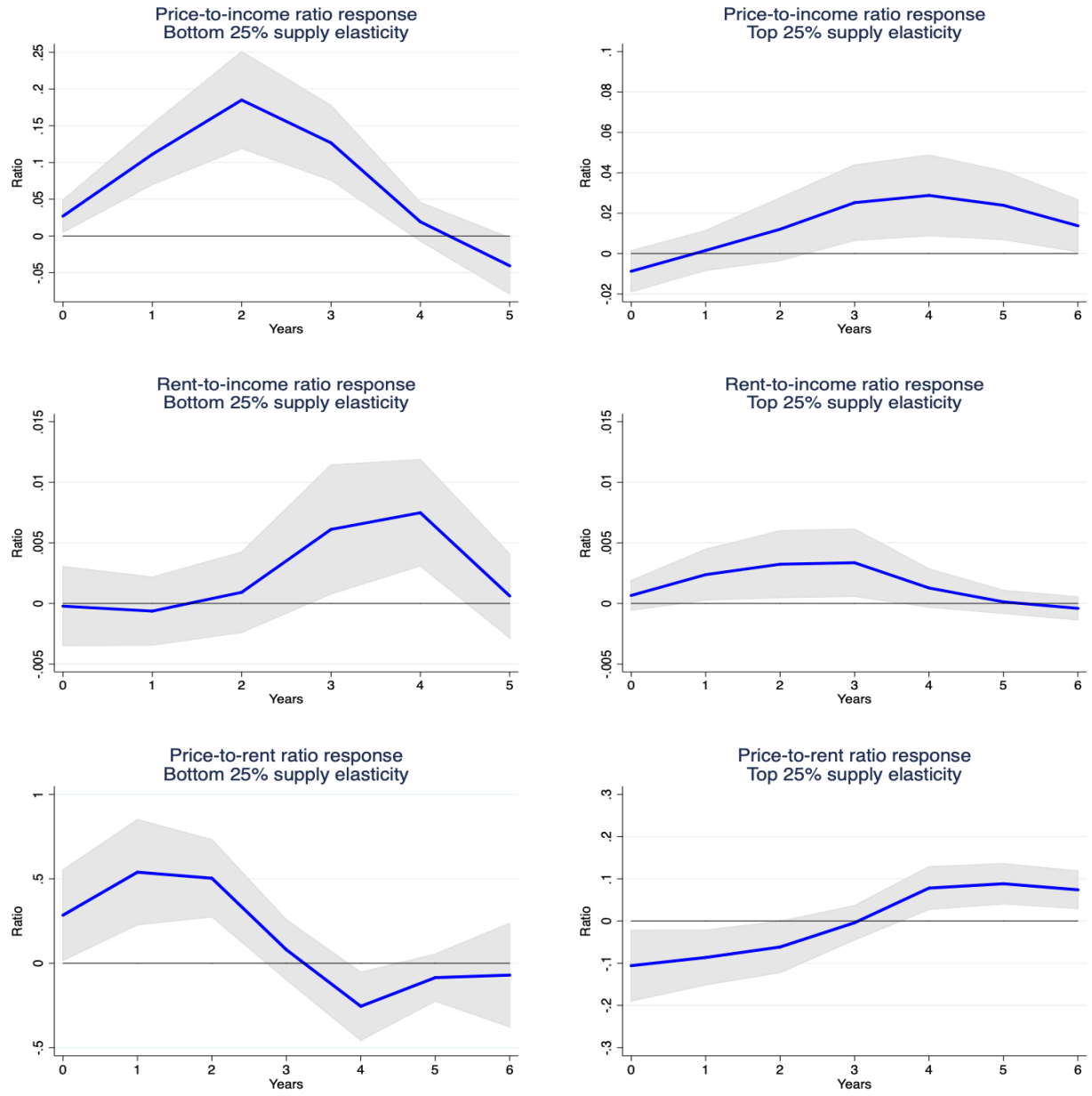

Figure 8. Dynamics of housing affordability after investors' purchases and supply elasticity. The figure plots the estimates from sequential regressions of (a) price-to-income, (b) rent-to-income, and (c) price-to-rent ratio for MSAs at the bottom and top quartiles of the supply elasticity distribution. Housing supply elasticity comes from Saiz (2010). The bottom quartile of the supply elasticity is 1.56, and the top quartile is 2.89 in our sample. The bottom quartile has on average 4.2 price-to-income, 0.33 rent-to-income and 13.0 price-to-rent ratio, over 2009-2017. The top quartile has on average 2.5 price-to-income, 0.27 rent-to-income and 9.4 price-to-rent ratio, over the same period. The shaded areas show the 90% confidence interval.

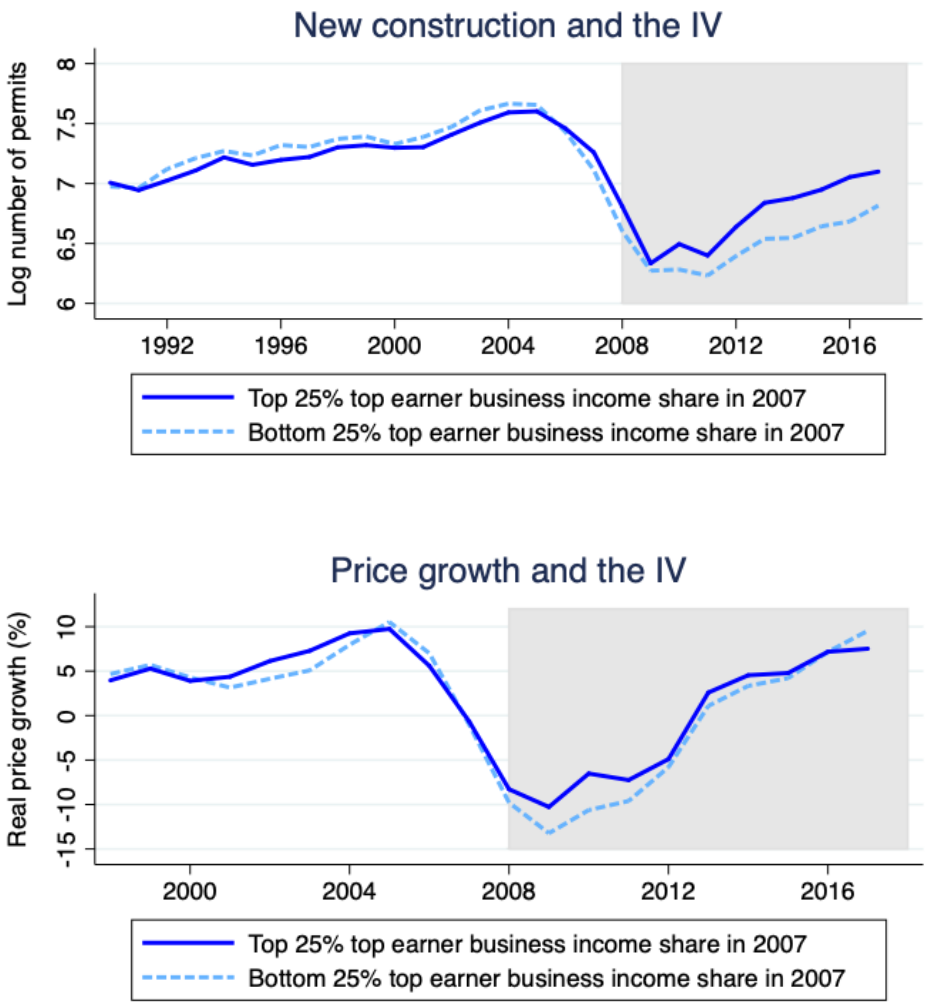

Figure 9. Parallel trends. The top figure plots the time series of the log number of new building permits for MSAs ranking in the top and bottom 25% of exposure to the instrumental variable (that is 2007 top earners' average share of business income over total income in a MSA). The bottom figure plots the same MSAs but for the bottom-tier real price growth The gray shaded area shows the period from the introduction of the Fed's QE in 2008 onwards.

Tables

Table 1. Summary statistics

Panel A - MSA level

	Obs	Mean	SD	Min	Max
Investors' share of purchases (%)	332	12.37	7.78	3.10	41.26
Top tier price growth (%)	328	0.43	1.55	-4.26	6.45
Mid-tier price growth (%)	332	0.47	1.77	-5.15	5.96
Bottom tier price growth (%)	296	0.17	2.52	-8.97	7.04
Top tier price-to-income ratio	328	3.09	1.22	1.40	9.50
Mid tier price-to-income ratio	332	4.76	2.40	1.50	16.98
Bottom tier price-to-income ratio	296	8.48	5.48	1.05	38.68
Log number of building permits all properties	332	6.49	1.27	2.33	10.33
Log number of building permits single-unit	332	6.44	1.28	2.24	10.31
Log number of building permits 2–4 units	330	2.25	1.23	0	6.73
Log number of building permits 5+ units	327	2.41	1.27	0	6.43
Top earner business income share$_{2007}$(%)	332	2.77	0.94	1.03	9.09

Panel B - Panel data

Investors' share of purchases (%)	2,997	11.50	8.40	0.65	75.95
Top tier price growth (%)	2,853	0.46	5.61	-24.92	28.41
Mid tier price growth (%)	2,901	0.47	6.67	-25.51	36.47
Bottom tier price growth (%)	2,610	0.13	9.87	-53.03	34.09
Rent growth (%)	2,583	0.52	6.12	-35.07	49.65
Price-to-income ratio of median household	2,849	3.24	1.27	1.12	9.97
Rent-to-income ratio of median household	2,583	0.29	0.05	0.14	0.61
Log number of building permits all properties	2,997	6.46	1.36	1.10	10.58
Log number of homeowner vacancies	2,554	7.57	1.13	3.14	10.96
Lagged population growth (%)	2,994	0.71	0.90	-4.45	7.99
Lagged median household income growth (%)	2,853	1.41	2.61	-7.98	11.01
Lagged unemployment rate change (%)	2,997	0.04	1.56	-4.54	9.29
Top earner business income share$_{07}$(%)×CD rate growth$_{t-1}$	2,997	-0.57	0.76	-4.98	1.58

The top panel presents summary statistics of the key variables at MSA level, and the bottom panel at MSA-year level, in 2009-2017. Prices and rents are inflation adjusted to reflect 2012 dollars. Detailed description of the variables and data sources is in Appendix A.

Table 2. First stage: Investors' share and the instrumental variable

	Investors' share of purchases$_{m,09-17}$
Top earner business income share$_{m,07}$	1.441***
	(0.327)
MSA-level controls	Yes
State dummies	Yes
R-squared	0.689
Observations	332

Heteroskedasticity robust standard errors are in parentheses. The controls are the population growth, income growth, unemployment rate change and real housing price growth over the periods 2000-2006 and 2006-2007, and the log number of building permits in 2007. Each observation is an MSA. ***significant at the 1% level.

Table 3. Housing price growth and affordability by price tier

Price growth$_{m,09-17}$				
	Mid Tier		Bottom Tier	Top Tier
Investors' share$_{m,09-17}$	0.034**	0.243***	0.302***	0.180**
	(0.015)	(0.083)	(0.100)	(0.070)
Estimation	OLS	IV	IV	IV
1st stage F-test excluded instruments		19.430	19.457	19.969
Underidentification test p-value		0.000	0.000	0.000
Observations	332	332	296	328
			Standardized	
Investors' share$_{m,09-17}$		0.827***	0.909***	0.768**
		(0.313)	(0.302)	(0.338)
Observations		293	293	293
	Price-to-income ratio$_{m,09-17}$			
Investors' share$_{m,09-17}$	0.108***	0.539***	1.538***	0.304***
	(0.021)	(0.161)	(0.375)	(0.096)
Estimation	OLS	IV	IV	IV
Observations	332	332	296	328
			Standardized	
Investors' share$_{m,09-17}$		1.679***	2.108***	1.884***
		(0.490)	(0.509)	(0.585)
Observations		293	293	293

Heteroskedasticity robust standard errors are in parentheses. Prices are inflation adjusted. Bottom tier refers to the 17th percentile, and top tier to the 83rd percentile of the housing prices and individual income in each MSA. The standardized results show the estimated effects of the standardized independent variable on the standardized dependent variables, for the sample of MSAs for which we have price series for all price tiers. All models include state dummies and MSA-level controls: population growth, income growth, unemployment rate change and real housing price growth over the periods 2000-2006 and 2006-2007, and the log number of building permits in 2007. Table 2 contains the first stage of the IV regression. The instrument for the investors' share of purchases is the average share of business income over total income of the top earners in MSA m in the year 2007. The weak identification F statistic is the Kleibergen and Paap Wald F statistic. The underidentification test is from Kleibergen and Paap (2006). Each observation is an MSA. ***significant at the 1% level; **significant at the 5% level.

Table 4. Affordability results excluding top MSAs

	Price growth$_{m,09-17}$			Price-to-income ratio$_{m,09-17}$		
	Bottom Tier	Mid Tier	Top Tier	Bottom Tier	Mid Tier	Top Tier
Sample without top 20 MSAs						
Investors' share$_{m,09-17}$	0.313***	0.252***	0.188**	1.728***	0.579***	0.342***
	(0.112)	(0.089)	(0.076)	(0.429)	(0.175)	(0.101)
F-test of excluded instruments	16.782	17.205	17.734	16.782	17.205	17.734
Underidentification p-value	0.000	0.000	0.000	0.000	0.000	0.000
Observations	276	312	308	276	312	308
Sample without top 37 MSAs						
Investors' share$_{m,09-17}$	0.331***	0.251***	0.178**	1.694***	0.570***	0.330***
	(0.108)	(0.089)	(0.079)	(0.435)	(0.166)	(0.090)
F-test of excluding instruments	16.535	17.163	17.724	16.535	17.163	17.724
Underidentification test p-value	0.000	0.000	0.000	0.000	0.000	0.000
Observations	259	295	291	259	295	291

Heteroskedasticity robust standard errors are in parentheses. Top MSAs are the ones with the largest dollar purchases by top 1% investors. These include the 20 largest investors in single-family rentals, and the apartment REITs in the S&P 500 Real Estate Sector. Prices are inflation adjusted. All models include state dummies, MSA-level controls and the instrumental variable as in Table 3. The weak identification F statistic is the Kleibergen and Paap Wald F statistic. The underidentification test is from Kleibergen and Paap (2006). Each observation is an MSA. ***significant at the 1% level; **significant at the 5% level.

Table 5. Housing construction by property type

	Log number of permits$_{m,09-17}$			
	All	Single-unit	2-4 units	5+ units
Investors' share$_{m,09-17}$	0.051***	0.047**	0.107*	0.152***
	(0.020)	(0.019)	(0.055)	(0.046)
Estimation	IV	IV	IV	IV
1st stage F-test excluded instruments	19.430	19.430	19.707	19.453
Underidentification test p-value	0.000	0.000	0.000	0.000
Observations	332	332	330	327

Heteroskedasticity robust standard errors are in parentheses. Single-unit refers to permits for the construction of single-unit properties, 2-4 units refers to permits for the construction of properties that have between 2 and 4 units, and 5+ units refers to permits for the construction of properties of 5 units or more. All models include state dummies, MSA-level controls and the instrumental variable as in Table 3. The weak identification F statistic is the Kleibergen and Paap Wald F statistic. The underidentification test is from Kleibergen and Paap (2006). Each observation is an MSA. ***significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 6. Placebo: Housing price growth and investors' share pre-crisis

	Price growth$_{m,[t_1,t_2]}$		
$[t_1, t_2]$	2000-2006	2001-2006	2000-2005
Investors' share of purchases$_{m,[t_1,t_2]}$	0.027	0.870	-0.036
	(0.807)	(1.680)	(2.238)
Estimation	IV	IV	IV
MSA-level controls	Yes	Yes	Yes
State dummies	Yes	Yes	Yes
Observations	307	303	306

Heteroskedasticity robust standard errors are in parentheses. Prices are inflation adjusted. The controls are the population growth, income growth, unemployment rate change and real housing price growth over the periods 1991-1997 and 1997-1998, and the log number of construction unit permits in 1998. The instrument for the investors' share of purchases is the average share of business income over total income of the top earners in MSA m in the year 2007. Each observation is an MSA.

Table 7. Estimation including additional local economic drivers

	Price growth$_{m,09-17}$				
Investors' share$_{m,09-17}$	0.255***	0.248***	0.220***	0.243***	0.241***
	(0.084)	(0.085)	(0.084)	(0.084)	(0.090)
Unempl. rate change$_{m,09-17}$	-3.245**				-2.787*
	(1.307)				(1.468)
Labor force partic. growth$_{m,09-17}$		-0.035			-0.129
		(0.217)			(0.240)
Real per cap. GDP growth$_{m,09-17}$			0.194		0.180
			(0.133)		(0.157)
Per cap. wage growth$_{m,09-17}$				-0.010	-0.187
				(0.197)	(0.219)
First stage F-test	18.937	19.326	17.686	19.522	16.868
Underidentification test p-value	0.000	0.000	0.000	0.000	0.000
Observations	332	331	332	332	331

Heteroskedasticity robust standard errors are in parentheses. Unemployment rate change$_{m,09-17}$ denotes the average unemployment rate change in MSA m over 2009-2017. Labor force participation growth$_{m,09-17}$, real per capita GDP growth$_{m,09-17}$ and per capita wage growth$_{m,09-17}$ denote the average annual growth rate of those variables in MSA m over 2009-2017. Prices are inflation adjusted. The specifications include MSA-level controls, state dummies and the instrumental variable as in Table 3. The underidentification test is that of Kleibergen and Paap (2006) and the F statistic is the Kleibergen and Paap Wald F statistic. Each observation is an MSA. ***significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 8. Estimation including credit condition controls

	Price growth$_{m,09-17}$	
Investors' share$_{m,09-17}$	0.222***	0.238***
	(0.077)	(0.081)
Mortgage application denial rate$_{m,09-17}$	-0.045	
	(0.040)	
Tested lenders' share$_{m,2008}$		-0.005
		(0.007)
First stage F-test	20.400	19.653
Underidentification test p-value	0.000	0.000
Observations	332	332

Heteroskedasticity robust standard errors are in parentheses. Mortgage application denial rate$_{m,09-17}$ is the average share of mortgage applications that were denied annually in MSA m over 2009-2017. Tested lenders' share$_{m,2008}$ is the 2008 deposit share of lenders in MSA m that underwent a stress test between 2011 and 2017. Prices are inflation adjusted. The specifications include MSA-level controls, state dummies and the instrumental variable as in Table 3. The underidentification test is that of Kleibergen and Paap (2006) and the F statistic is the Kleibergen and Paap Wald F statistic. Each observation is an MSA. ***significant at the 1% level.

Table 9. Estimation controlling for labor demand shifts by industry

	Price growth$_{m,09-17}$					
	coef.	s.e.	coef.	s.e.	coef.	s.e.
Investors' share of purchases$_{m,09-17}$	0.236***	(0.087)	0.221***	(0.082)	0.224***	(0.085)
Employment growth by industry$_{m,08-17}$						
Health Care & Social Assistance	-0.001	(0.003)	-0.004	(0.004)	-0.004	(0.004)
Retail Trade	0.296**	(0.138)	0.274**	(0.132)	0.275**	(0.137)
Accommodation & Food Services	0.029	(0.094)	0.018	(0.091)	0.009	(0.094)
Manufacturing	-0.001	(0.006)	-0.008*	(0.005)	-0.009*	(0.005)
Professional, Scientific, Tech. Services	0.003	(0.003)	0.003	(0.002)	0.003	(0.002)
Administrative, Support, Waste Mgmt.	-0.001*	(0.000)	-0.001**	(0.000)	-0.001**	(0.000)
Finance & Insurance	0.002	(0.002)	0.002	(0.002)	0.002	(0.002)
Wholesale Trade			0.032	(0.034)	0.031	(0.036)
Other Services			0.096**	(0.042)	0.094**	(0.043)
Transportation & Warehousing			0.025**	(0.012)	0.023**	(0.012)
Information			0.004	(0.003)	0.005	(0.003)
Educational Services					-0.000	(0.001)
Management of Companies					-0.001	(0.001)
Real Estate, Rental & Leasing					0.003	(0.003)
Arts, Entertainment & Recreation					-0.000	(0.001)
1st stage F-test of excluded instruments	16.993		18.165		17.050	
Underidentification test p-value	0.000		0.000		0.000	
Observations	332		332		330	

Heteroskedasticity robust standard errors are in parentheses. The specifications control for the average annual growth in the number of employees in various industries - based on the North American Industry Classification System (NAICS) 2 digit sector codes - that predominate the labor market of MSAs, over 2008-2017. Prices are inflation adjusted. The specifications include MSA-level controls, state dummies and the instrumental variable as in Table 3. The underidentification test is that of Kleibergen and Paap (2006) and the F statistic is the Kleibergen and Paap Wald F statistic. Each observation is an MSA. ***significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 10. The instrumental variable and its predictors

	Top earner business income share$_{m,07}$	
Median age$_{m,07}$	0.030*	0.011
	(0.016)	(0.017)
Immigrants as % of population$_{m,07}$	0.032***	-0.001
	(0.008)	(0.010)
Income tax rate for top earners$_{m,07}$	0.055***	0.053***
	(0.018)	(0.016)
Entrepreneurship rank$_{m,07}$	-0.0001	0.001
	(0.003)	(0.003)
Natural amenity index$_{m,07}$		0.121***
		(0.022)
R-squared	0.113	0.223
Observations	280	277

Heteroskedasticity robust standard errors are in parentheses. The outcome variable is our instrument for the investors' share of purchases: the average share of business income over total income of the top earners in MSA m in the year 2007. Each observation is an MSA.

Online Appendix (NOT FOR PUBLICATION)

A Detailed Description of Database

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

Investors' purchases

The investors' data come from the Zillow Transaction and Assessment Dataset (ZTRAX), a large raw database of U.S. deeds data. The transactions database of ZTRAX contains all property ownership transfers that are documented in the County deeds. Each record contains the date of the transfer, the address of the property, the type of the property, the sale price, and the names of the buyer and seller. We keep transactions between January 1^{st}, 2000 and December 31^{st}, 2017. We restrict the data to ownership transfers, dropping observations that refer exclusively to mortgages or foreclosures.[20] We drop transactions with deed type "Life Estate", since this is not an immediate transfer of ownership. We also drop transactions that had "Cancellation" in the deed type. We restrict the data to residential property transfers based on the ZTRAX property land use standard codes, which include both single-family and multi-family properties. Table A14 contains the classification of the property land use standard codes in single-family and multi-family from ZTRAX. This amounts to 139 million transactions nationally. We then drop transactions with purchase price missing or smaller than \$10,000, a common practice with deeds data (Bernstein, Gustafson and Lewis 2019; Stroebel 2016). This leaves 85 million transactions. Table A15 describes step by step the construction of the database of transactions with the investors' classification.

With the previous cleaning criterion, most of the transactions are dropped in the non-disclosure states. These states or counties do not require that the sale price is submitted to the county office. Specifically, all transactions are dropped in five non-disclosure states: Mississippi, Missouri, Montana, Utah and Wyoming. We keep in our data seven non-disclosure states, with a total of 28 MSAs, in which some of the transactions record sales price. We drop from our final dataset MSA-years that have fewer than 200 transactions, to avoid large outlier values, due to very few observations. The final dataset contains the following MSAs in non-disclosure states: Anchorage, Alaska; Boise City, Idaho; Alexandria, Baton Rouge, Hammond, Houma-Thibodaux, Lafayette, Lake Charles, Monroe, New Orleans-Metairie and Shreveport-

[20] The mortgage and foreclosure deeds have a separate corresponding deed for the ownership transfer.

Bossier City, Louisiana; Kansas City and Wichita, Kansas; Albuquerque, New Mexico; Bismarck and Fargo, North Dakota; Amarillo, Austin-Round Rock, Brownsville-Harlingen, Corpus Christi, Dallas-Plano-Irving, El Paso, Fort Worth-Arlington, Houston-The Woodlands-Sugar Land, Killeen-Temple, Lubbock, McAllen-Edinburg-Mission and San Antonio-New Braunfels, Texas. Additional results, not reported here, contain our baseline cross-sectional and dynamic analyses, dropping completely all non-disclosure MSAs. The results of both analyses hold with the same significance and even stronger results for the relevance tests for our instrumental variable.

To identify *institutional* or *corporate investors*, we first use the ZTRAX classification of buyer names into individual and non-individual names. The non-individual names frequently end with the words "LLC", "LP", "INC", "TRUST", "CORPORATION", "PARTNERS", but they also contain entity names without the description in the end of the name.[21] Thorough inspection of the data confirms that the classification by ZTRAX of individual and non-individual names is as expected, with very minimal (human) errors. Our investors' identifier contains the deeds where the buyer has a non-individual name. 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, hospitals, universities (not when is university housing), churches, airports, and the state, names of the county, city and 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 further increase the accuracy of the largest investors' classification we collect from industry reports and news reports the names of the top 20 institutional investors in the single-family rental market. For example Amherst Capital's 2018 market commentary report[22] provides a comprehensive list of the top 20 single-family rental institutions and the number of homes owned based on their calculations. We also collect the names of the residential real estate companies that belong to the S&P 500 Real Estate Index, most of which are apartment REITs. We then search for the names of these top investors and their subsidiaries in the ZTRAX database and ensure they are classified as investors. We use public SEC filings and other business websites

[21] For example "Invitation Homes" and "Invitation Homes LP" are both included as non-individual names.

[22] Amherst Capital report is retrieved from https://www.amherstcapital.com/documents/20649/22737/Amherst+Capital+Market+Commentary+-+April+2018+vF/f06bd51a-44c7-4f8f-87e3-ca8d795bf42a Last visited: 03-05-2019.

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 and multi-family investors.

We calculate the market share of investors as the *dollar value* of investors' purchases divided by the dollar value of all purchases. Using the dollar value, accounts correctly for purchases of buildings with multiple units.

Alternatively, we use the *number of units*, instead of the dollar value. The number of units is coded by ZTRAX, in the tax assessment dataset, which we merge with the transactions dataset, using the RowID unique identifier. We use the property type code (PropertyLandUseStndCode) to fill in the missing number of units. Specifically, we fill in the number of units as 2 if the number of units is missing and the property type is duplex or multifamily dwelling (generic any combination 2+). We fill in the number of units as 3 for triplex, 4 for quadruplex, and 5 for apartment buildings (5+ units) or court apartments (5+ units). We fill in the number of units as 100 for apartment buildings (100+ units). With this criterion, when the number of units is missing we assign the lower bound of the number of units to the property, inferred by the qualitative description. For the rest of the multi-family property types and all the types we classify as single-family in Table A14 that do not specify number of units, we assign 1 unit. We double-check with the sales price and confirm that these refer to single-unit purchases.

Finally, we use the crosswalk file from Census Bureau to match the County FIPS codes in ZTRAX to the Census Bureau MSA's 2017 core based statistical area (CBSA) code. For submetro areas of the largest MSAs, we use the CBSA division code. In total we match 411 CBSAs and divisions in the data.

Housing prices, rents and supply elasticity

Our price and rent data at MSA-level from 1999 through 2017 come from Zillow. To measure housing prices, we use the Metro Zillow Home Value Index (ZHVI). The ZHVI measures the median monthly price for each MSA and has units of nominal dollars per month. Zillow imputes this price based on a proprietary machine learning model taking into account the specific characteristics of each home and recent sale listings for homes with similar characteristics. The median price is computed across all homes in an MSA, not only those that are currently for sale. Thus, unlike pure repeat-listing indices, the ZHVI is not biased by the current composition of for-sale properties. To measure housing prices specifically for single-family homes, we use the ZHVI Single-Family Homes Time Series. To measure the price of top tier and bottom tier homes we use the Zillow's Top Tier Index and Bottom Tier Index, which measure the median house price among homes in the top third and bottom third of the price distribution within

an MSA respectively. To measure rents, we use the Metro Zillow Rent Index (ZRI). The ZRI measures the median quarterly rent for each MSA and has units of nominal dollars per month. Zillow imputes this rent using an analogous methodology to ZHVI. Importantly, the ZRI does not impute a property's rent from its price. To convert the prices and rents to annual, we take the last value of each year. Housing price growth is the percentage growth of housing prices from year $t-1$ to year t. Housing rent growth is the percentage growth of housing rents from year $t-1$ to year t.

The housing supply elasticities were originally estimated by Saiz (2010). The elasticities are based on the amount of developable land in the U.S. MSAs, which is calculated based on satellite-generated geographical data. We use the dataset provided by Favara and Imbs (2015) as our source of elasticity data.[23] The original data are at the MSA level (CBSA 2003 codes), and cover 275 MSAs. We crosswalk these to our 2017 CBSA and CBSA division codes.

Construction and vacancy data

Data on construction permits come from the Census Bureau's annual Residential Building Permits Survey. Statistics on construction authorized by building permits are based upon reports submitted by local building permit officials in response to a mail survey. When a report is not received, missing residential data are either obtained from the Survey of Use of Permits (SUP) or imputed. The SUP is used to collect information on housing starts. All other missing data are imputed. The imputations are based on the assumption that the ratio of current year authorizations to those of a year ago should be the same for both respondents and nonrespondents.

Our construction data cover the years 2000 to 2017 and they are collected initially at the county level. We then use the crosswalk file from Census Bureau to match the County FIPS codes to the Census Bureau 2017 core based statistical area (CBSA) and CBSA division codes. Then we aggregate the number of construction permits at the CBSA level. The permits are split into 1-unit, 2-units, 3-4 units and 5+ units, and they count the number of new buildings authorized. For our main construction variable we add up all the permits together, since our analysis includes the total housing market. The MSA-level data cover all the 411 CBSA codes.

Vacancy data come from the American Community Survey One-Year Estimates. Data are available annually and they cover 311 MSAs over the 2005-2017 period. We start from the original data at the county level: number of vacant housing units for homeowners and number

[23] The AER site from which we obtained the data is: https://www.aeaweb.org/articles?id=10.1257/aer.20121416, and the specific dataset is "hp_dereg_controls".

of total units for homeowners. We then crosswalk to the 2017 CBSA codes and CBSA division codes and sum the number of households in the counties within the MSAs. Starting from county-level data results in more accurate MSA values for the most recent CBSA codes. Owner vacancy rate is the share of the number of vacant housing units for homeowners over the total housing units for homeowners.

Tax report data

The main data source to construct our instruments comes from the Internal Revenue Services (IRS), in particular, the Statistics of Income (SOI). This dataset provides zip code data on administrative records of individual tax returns. The data excludes zip codes with less than 100 returns. Detailed description of the instruments is included in Appendix B.

Control variables

We also rely on the following data sources to get data at the county-year level and then aggregate to MSA-year level using the 2017 CBSA and CBSA division codes:

- Population: U.S. Census Bureau, from 1990 to 2017.
- Median Income: Zillow Median Household Income dataset, from 1990 to 2017.
- Unemployment and labor force participation: Bureau of Labor Statistics, from 1990 to 2017.
- Median age: American Community Survey One-Year Estimates, Census Bureau. The data only cover the 2005-2017 period. The data come in discrete age intervals that are 5 years apart. Based on the number of people in each age interval we find the interval that contains the median age, and take as the median age the midpoint of this interval.
- Employment by industry: County Business Patterns (CBP) dataset, from 2007 to 2017.
- Gross Domestic Product and wages: U.S. Department of Commerce's Bureau of Economic Analysis (BEA), from 2008 to 2017.
- Natural Amenities Scale: U.S. Department of Agriculture. The scale is constructed by combining six measures of climate, topography, and water area that reflect environmental qualities. These measures are warm winter, winter sun, temperate summer, low summer humidity, topographic variation, and water area.

These additional controls come from the following data sources:

- Migration: American Community Survey 5-year estimates, Census Bureau, MSA-level in 2007.
- Income tax rate: Tax Foundation, the top marginal tax rate for an individual, State-level in 2007.
- Entrepreneurship rank: CNBC America's top states for business in 2007. This index provides a ranking of 50 States based on 40 different measures of competitiveness from publicly available data.

To summarize, there are 332 MSAs with the full set of average housing variables and investors' market share for the years 2009-2017, control variables beginning in 2000, and tax-returns for the year 2007.

B Detailed Description of the Instrumental Variable

Our instrument approximates the average individual's tax returns by the zip code returns of a specific adjusted gross income (AGI) group. Since the Statistics of Income (SOI) dataset from the IRS does not provide returns at the individual level, the zip code AGI group level is the closest approximation to the average individual of each group within the zip code. AGI is defined as the total income minus adjustments to the income, which might be subject to change each year. The dataset splits the returns into six income groups. We specifically focus on the returns of the top two high earnings groups, which include people with annual AGI above $100,000.

Our instrument is the share of business income which measures the local attitude towards investment. Next, we describe in detail how we construct this instrument.

Share of business income

The share of business income instrument is concerned with the component of earnings associated with net business income. With the implementation of the QE housing becomes an attractive investment. High earners with high business income in each MSA are likely to be more knowledgeable about investments. They are more likely to pursue investments in general, and investments in residential real estate in particular.

To construct the instrument we calculate the average share of net business income of top earners in 2007 at zip code level as:

$$b_{z,2007} = \sum_{g=5}^{6} \mu_g \frac{\text{Net business income (\$)}_g}{\text{Adjusted gross income (\$)}_g},$$

where z denotes the zip code and $g \in \{5, 6\}$, denotes the AGI group. Group 5 consists of returns with AGI between \$100,000 and \$200,000, and group 6 consists of returns with AGI above \$200,000. The weight μ_g weights by the number of returns of each group. $\mu_g = N_g/(N_5 + N_6)$, where N represents the number of returns. All values refer to the 2007 returns.

We calculate the average share of business income of top earners in 2007 at the MSA level as:

$$b_{m,2007} = \sum_{z \in m} \omega_z k_z b_z,$$

where m denotes the MSA. k_z is the share of the zip code population that belongs to the MSA. This share comes from the Department of Housing and Urban Development (HUD) zip-CBSA and zip-CBSA division crosswalk files. k_z is one for most of the zip codes. ω_z weights by the number of returns of each zip code within the MSA: $\mu_z = N_z / \sum_{z \in m} N_z$.Our instrument $b_{m,2007}$ is used in the cross-sectional regression (1) to instrument for the average share of investors in MSA m, using a 2-stage least square estimation methodology.

For our dynamic analysis that uses a panel specification, we use the panel version of the instrument. The time-varying instrument captures the exposure of an MSA to the QE over time. We construct the time-varying instrument as follows:

$$b^p_{m,t} = b_{m,2007} \times CD_{t-1},$$

where CD_t is the growth in the one-year certificate of deposits rate from year $t-1$ to t. In our panel data t ranges from 2009 to 2017. The investors' share is used with one year lag in the panel specification (2) .

Having the business income share fixed in 2007, ensures that the exposure to the QE is predetermined, and not affected by the housing market variables post 2008. Figure 1 plots the time series of an average one-year CD rate. CD_t is a national shock that is also unrelated to each of the local housing markets. This methodology constructs instruments that are likely to satisfy the exclusion restriction. Our multiple tests in Section 5 provide strong evidence in this direction.

C Altonji, Elder and Taber (2005) and Oster (2019) Test of Omitted Variable Bias

Altonji, Elder and Taber (2005) and Oster (2019) propose a test for omitted variable bias that uses the values of the coefficient of interest and R-squared in two different regressions: with and without control variables. We estimate an interval for the coefficient of interest and confirm that this interval does not contain zero. Note that this methodology is applied to OLS regressions, since the fundamental principle is obtaining the best model fit, as measured by the R-squared (see Mian and Sufi 2014 and Jenwittayaroje and Jiraporn 2019 for applications).

We compute the identified interval for the coefficient of the investors' share $[\beta, \beta^*]$, where β is the coefficient of the fully controlled model, and $\beta^* = \beta - (\beta_{uncontrolled} - \beta)\frac{R_{\max} - \tilde{R}}{\tilde{R} - R_{uncontrolled}}$. $\tilde{R}$ is the R-squared of the fully controlled model. $\beta_{uncontrolled}$ and $R_{uncontrolled}$ are the coefficient and R-squared of the basic model without controls, respectively. For $R_{\max}$ we use both definitions suggested by Oster (2019), $R_{\max} = 1.3\tilde{R}$ or $R_{\max} = 2.2\tilde{R}$.

We perform this test for four different specifications: (1) our baseline specification in Table 3, (2) the specification with additional controls for economic drivers in Table 7, (3) the specification with additional controls for credit supply in Table 8 and (4) the specification that includes additional controls for changes in industry employment in Table 9.

Table A9 shows the results of these tests. For $R_{\max} = 1.3\tilde{R}$, the identified intervals are: $[0.0319, 0.0343]$, $[0.0221, 0.0269]$, $[0.0325, 0.0348]$ and $[0.0240, 0.0283]$, for the above models. The identified intervals safely exclude zero. Thus, we can reject that the effect of the share of investors on housing prices is driven by omitted variables.

Extra Figures (NOT FOR PUBLICATION)

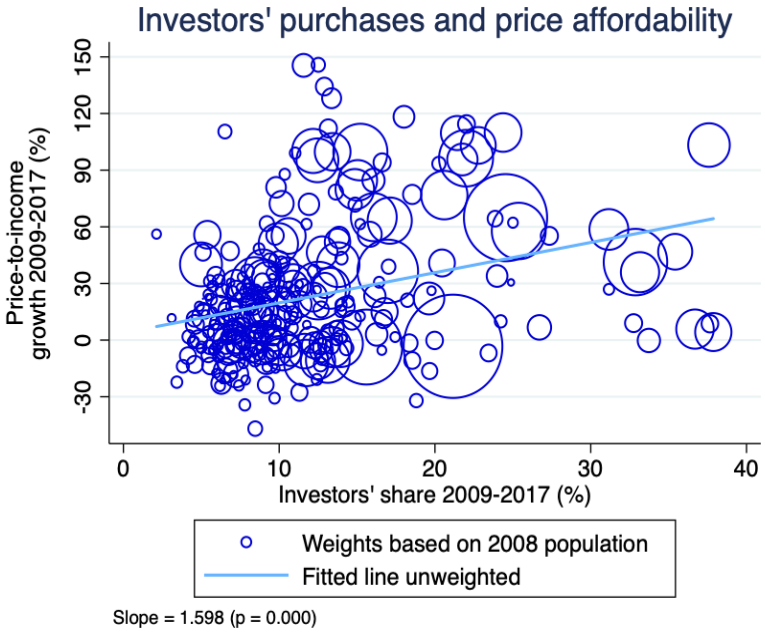

Figure A1. Affordability and investors in the U.S. The figure plots the average share of investors' purchases in the years 2009 to 2017 against the growth of the bottom-tier price-to-income ratio from 2009 to 2017 in the U.S. MSAs. Each circle represents an MSA, and the size of the circle is analogous to the MSA population in 2008.

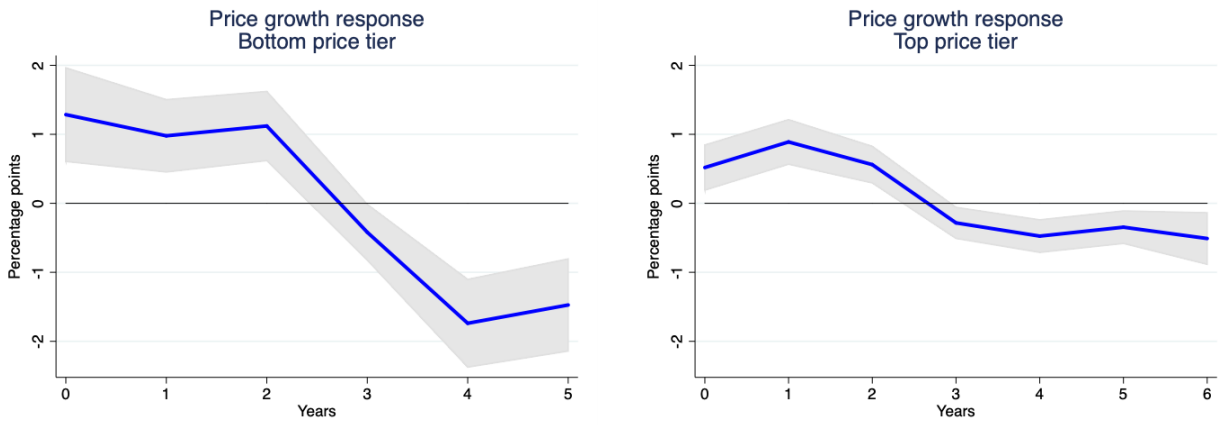

Figure A2. Dynamics of housing prices after investors' purchases by tier. The figure plots the estimates from sequential regressions of the real housing price growth on the instrumented past investors' share of purchases for top and bottom price-tier houses. Top tier houses are houses in the top third, and bottom tier houses are houses in the bottom third of the house value distribution within an MSA. We estimate the impulse responses for the full panel data from 2009 to 2017. The shaded areas show the 90% confidence interval.

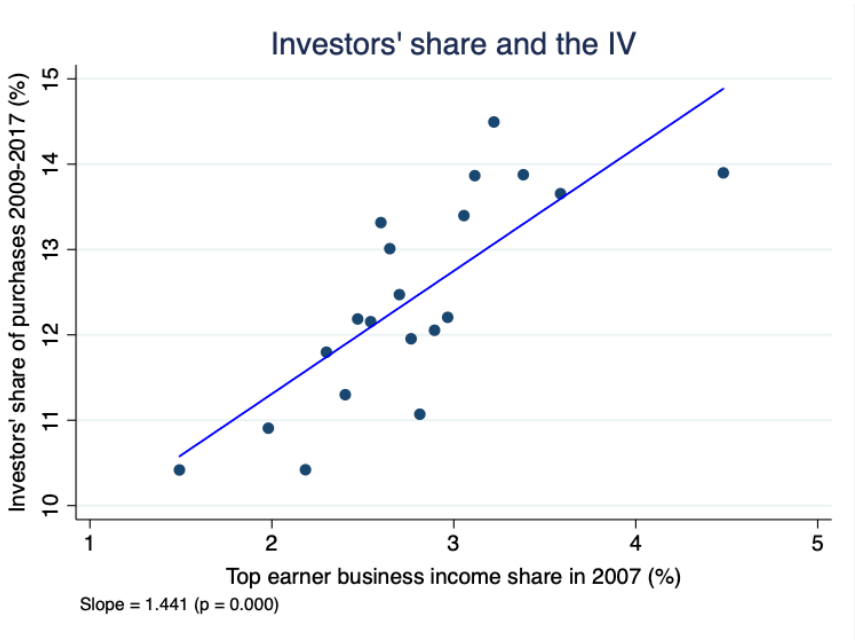

Figure A3. Investors' market share and the instrumental variable. This figure plots the average share of value of business income over total income of top earners in an MSA in 2007, against the 2009-2017 average market share of investors' purchases in each MSA. The top earners are the ones who reported adjusted gross income of 100,000 U.S. dollars or higher in their tax returns. The MSAs are binned by percentiles so that each point represents around 15 MSAs. The figure controls for the controls in the baseline specification in Table 3.

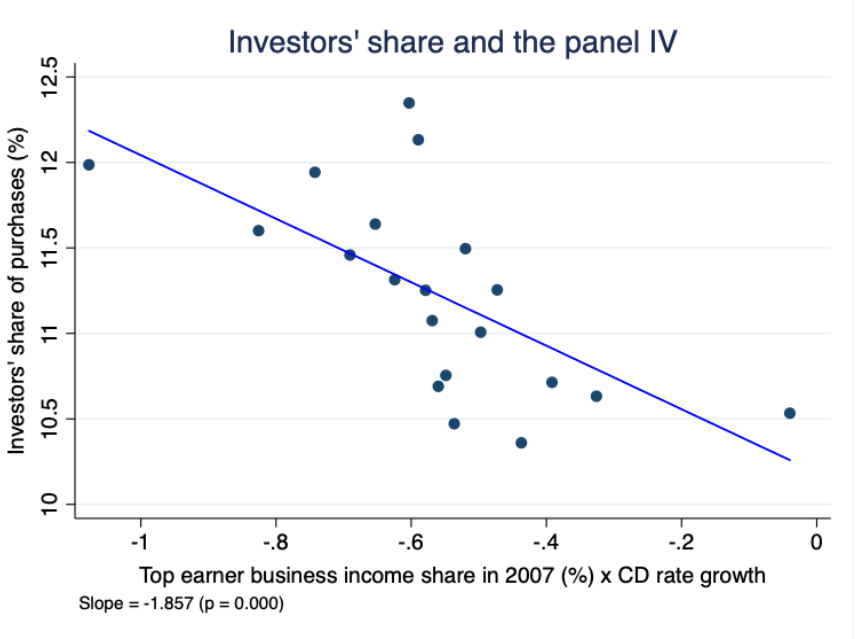

Figure A4. Investors' market share and the instrumental variable. This figure plots the share of value of business income over total income of top earners in an MSA in 2007 multiplied by the CD rate growth, against the market share of investors' purchases each year in each MSA. The top earners are the ones who reported adjusted gross income of 100,000 U.S. dollars or higher in their tax returns. The MSAs are binned by percentiles so that each point represents around 15 MSAs. The figure controls for the controls in the panel specification in Table A4.

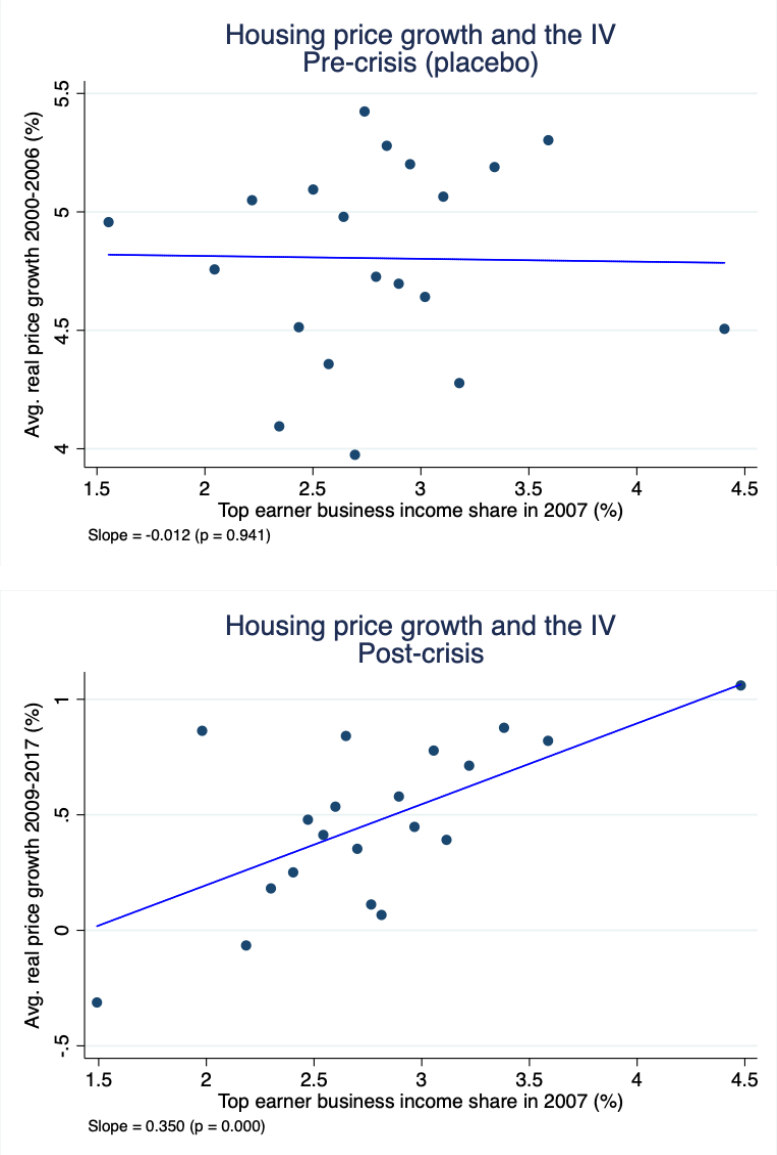

Figure A5. Housing price growth against the instrument for investors pre- and post-2008. The top panel plots the 2000–2006 average annual real housing price growth against the average share of business income over total income of top earners in each MSA in 2007. The bottom panel plots the 2009–2017 average annual real housing price growth against the same instrument. The top panel controls are the ones used in the placebo specification in Table 6. The bottom panel controls are the ones used in the baseline specification in Table 3. Figure A6 performs the same visual exercise for the panel version of the instrument.

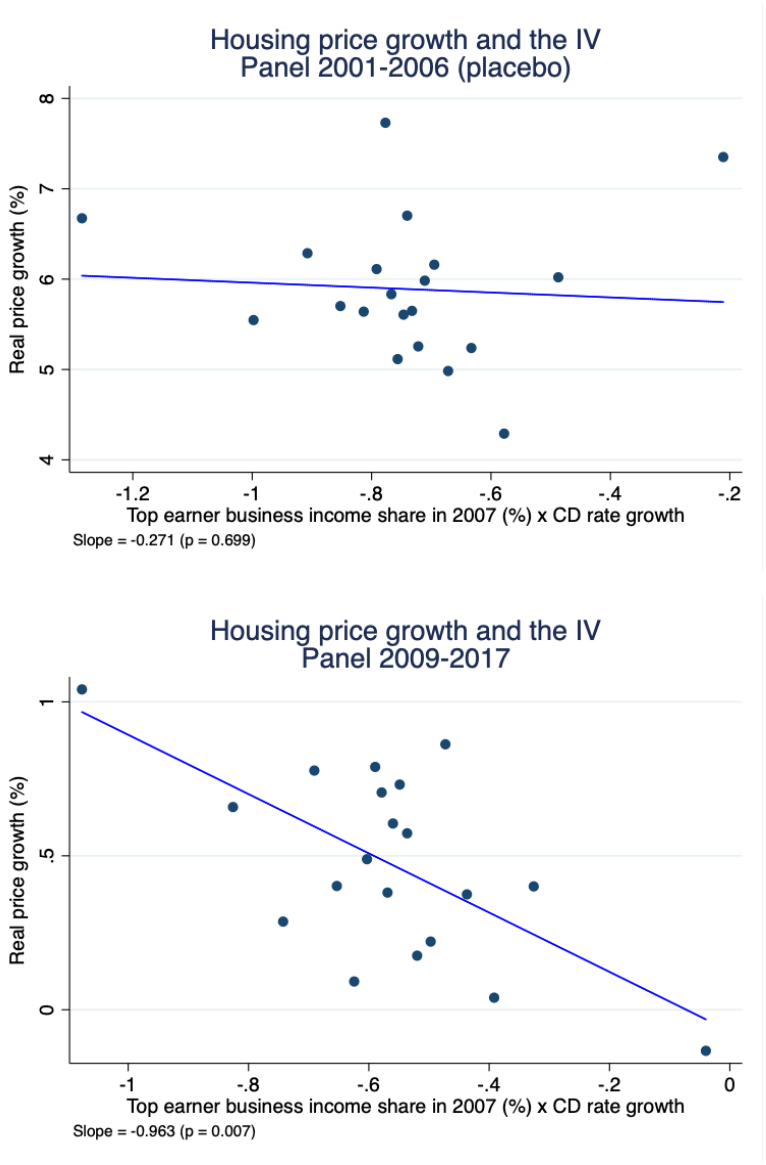

Figure A6. Pre- and post-2008 housing price growth against the panel instrument for investors. The top panel plots the annual real price growth over the 2001-2006 period against the panel instrument: the average share of business income over total income of top earners in each MSA in 2007 multiplied by the lagged CD rate growth. The bottom panel plots the annual real housing price growth over the 2009-2017 period against the same instrument. The top panel controls are the ones used in the placebo panel specification in Table A7. The bottom panel controls are the ones used in the panel specification in Table A4.

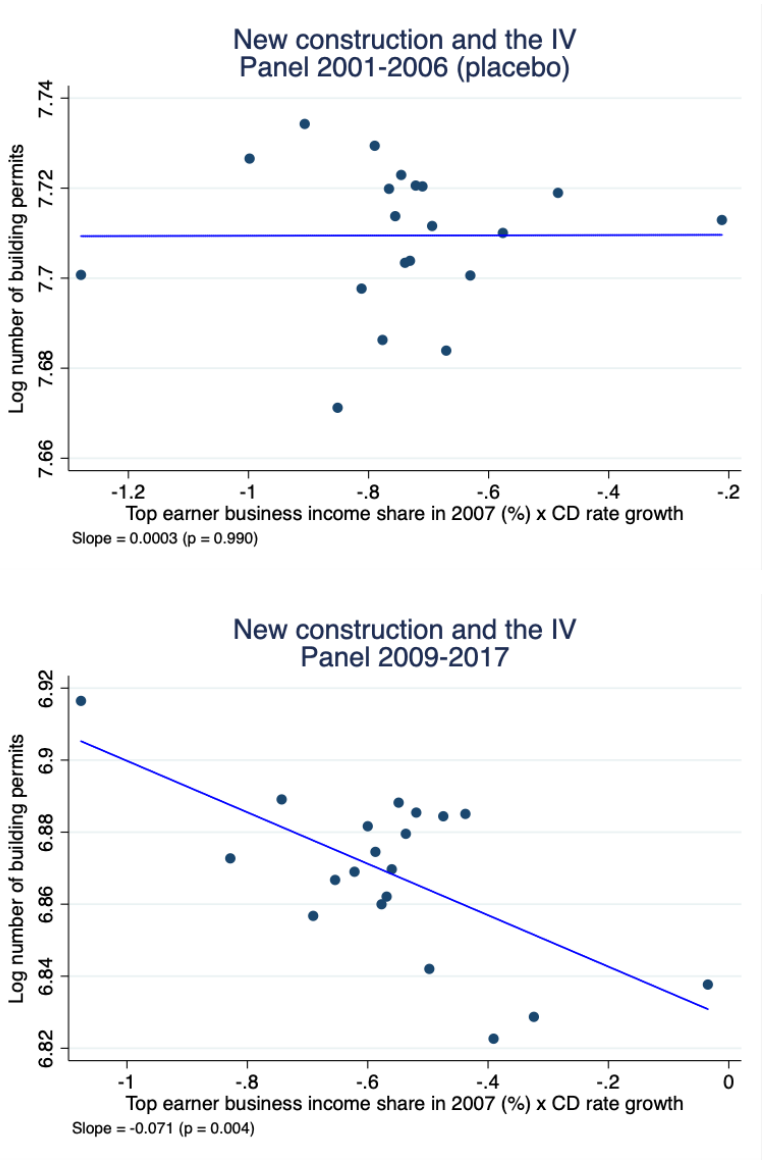

Figure A7. Pre- and post-2008 building permits against the panel instrument for investors. The top panel plots the log number of building permits over the 2001-2006 period against the panel instrument: the average share of business income over total income of top earners in each MSA in 2007 multiplied by the lagged CD rate growth. The bottom panel plots the log number of building permits over the 2009-2017 period against the same instrument. The top panel controls are the ones used in the placebo panel specification in Table A7. The bottom panel controls are the ones used in the panel specification in Table A4.

Extra Tables (NOT FOR PUBLICATION)

Table A1. County level results

	Bottom Tier	Mid Tier	Top Tier
	Price growth$_{m,09-17}$		
Investors' share$_{m,09-17}$	0.361***	0.264**	0.144**
	(0.136)	(0.125)	(0.069)
Observations	601	691	676
	Price-to-income ratio$_{m,09-17}$		
Investors' share$_{m,09-17}$	1.675***	0.659***	0.478***
	(0.566)	(0.222)	(0.096)
Observations	601	691	676

Heteroskedasticity robust standard errors are in parentheses. Bottom tier houses are houses in the bottom third, and top tier in the top third of the house value distribution within a County. Investors' share is the average annual market share of purchases by investors in County c over 2009-2017. All models include state dummies and county-level controls: population growth, income growth, unemployment rate change and real housing price growth over the periods 2000-2006 and 2006-2007, and the log number of construction unit permits in 2007. Price growth is inflation adjusted. The instrument for the investors' share of purchases is the average share of business income over total income of the top earners in County c in the year 2007. Each observation is a County. ***significant at the 1% level; **significant at the 5% level.

Table A2. Construction results excluding top MSAs

	All	Single-unit	2-4 units	5+ units
Sample without top 20 MSAs				
Investors' share$_{m,09-17}$	0.057***	0.053***	0.121**	0.164***
	(0.021)	(0.020)	(0.060)	(0.052)
Observations	312	312	310	307
Sample without top 37 MSAs				
Investors' share$_{m,09-17}$	0.058***	0.055***	0.093*	0.143***
	(0.020)	(0.019)	(0.051)	(0.049)
Observations	295	295	293	290

Heteroskedasticity robust standard errors are in parentheses. Top MSAs are the ones with the largest dollar purchases by top 1% investors. These include the 20 largest investors in single-family rentals, and the apartment REITs in the S&P 500 Real Estate Sector. Single-unit refers to permits for the construction of single-unit properties, 2-4 units refers to permits for the construction of properties that have between 2 and 4 units, and 5+ units refers to permits for the construction of properties of 5 units or more. All models include state dummies, MSA-level controls and the instrumental variable as in Table 3. The underidentification test is from Kleibergen and Paap (2006). Each observation is an MSA. ***significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table A3. First stage panel: Investors' share and the instrumental variable

	Investors' share$_{m,t-1}$
Top earner business income share$_{m,07}$×CD rate growth$_{t-2}$	-1.857***
	(0.368)
MSA-year controls	Yes
MSA fixed effects	Yes
Year fixed effects	Yes
R-squared	0.691
Observations	2,842

Standard errors clustered by MSA are in parentheses. The controls are the housing price growth, population growth, median income growth and unemployment rate change, all lagged by one year. The sample period is 2009-2017. Each observation is an MSA-year. ***significant at the 1% level.

Table A4. Housing price growth in response to investors' purchases

	Price growth$_{m,t+i}$						
	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$
Top price-tier							
Investors' share$_{m,t-1}$	0.52***	0.89***	0.56***	-0.28**	-0.48***	-0.35**	-0.51**
	(0.20)	(0.20)	(0.16)	(0.14)	(0.14)	(0.14)	(0.23)
Observations	2,804	2,492	2,180	1,868	1,556	1,243	932
Mid price-tier							
Investors' share$_{m,t-1}$	0.52***	0.86***	0.70***	-0.48***	-0.78***	-0.40***	-0.74***
	(0.20)	(0.24)	(0.20)	(0.15)	(0.18)	(0.14)	(0.28)
Observations	2,842	2,525	2,207	1,891	1,575	1,258	942
Bottom price-tier							
Investors' share$_{m,t-1}$	1.29***	0.98***	1.12***	-0.42*	-1.74***	-1.47***	-2.63**
	(0.41)	(0.32)	(0.31)	(0.25)	(0.39)	(0.41)	(1.02)
Observations	2,547	2,260	1,974	1,690	1,406	1,118	837

Standard errors clustered by MSA are in parentheses. i indicates the number of years forward for which the effect is estimated. All models include location and time fixed effects and controls: the lagged dependent variable, and population growth, median household income growth and unemployment rate change, all lagged by one year. Prices are inflation adjusted. The IV is the average share of business income over total income of the top earners in MSA m in 2007 multiplied by the lagged CD rate growth. The sample period is 2009-2017. Each observation is an MSA-year. The Kleibergen and Paap (2006) underidentification test has p-value of 0.001, and the Kleibergen and Paap Wald F statistic is 25.475 for the mid-tier market panel regression ($i=0$). Table A3 contains the first stage of the IV regression. Table A13 contains the dynamic results using alternative measures of the investors' presence. ***significant at 1%; **significant at 5%; *significant at 10%.

Table A5. Affordability measures in response to investors' purchases

	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$
	Price-to-income ratio$_{m,t+i}$						
Investors' share$_{m,t-1}$	0.03***	0.09***	0.14***	0.11***	0.04***	0.01	-0.01
	(0.01)	(0.02)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)
Observations	2,844	2,527	2,210	1,892	1,576	1,259	944
	Rent-to-income ratio$_{m,t+i}$						
Investors' share$_{m,t-1}$	0.06	0.10	0.22**	0.40***	0.30***	-0.02	-0.21*
	(0.10)	(0.09)	(0.10)	(0.14)	(0.09)	(0.06)	(0.12)
Observations	2,580	2,293	2,006	1,719	1,432	1,144	858

Standard errors clustered by MSA are in parentheses. i indicates the number of years forward for which the effect is estimated. All models include location and time fixed effects and controls: the lagged dependent variable, and population growth, median income growth and unemployment rate change, all lagged by one year. The IV is the average share of business income over total income of the top earners in MSA m in 2007 multiplied by the lagged CD rate growth. The sample period is 2009-2017. Each observation is an MSA-year. The investors' share is divided by 100 in the regressions of rent-to-income to adjust the coefficients. ***significant at 1%; **significant at 5%; *significant at 10%.

Table A6. Single-family properties

	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$
	Single-family price growth$_{m,t+i}$						
Investors' single-family share$_{m,t-1}$	0.61***	1.07***	0.86***	-0.64***	-0.98***	-0.47***	-1.10**
	(0.22)	(0.28)	(0.24)	(0.19)	(0.22)	(0.17)	(0.46)
Observations	2,830	2,514	2,197	1,882	1,567	1,250	936
	Price growth$_{m,t+i}$						
Investors' single-unit share$_{m,t-1}$	0.59***	1.05***	0.88***	-0.61***	-1.01***	-0.51***	-1.07**
	(0.21)	(0.27)	(0.24)	(0.18)	(0.22)	(0.18)	(0.43)
Observations	2,842	2,525	2,207	1,891	1,575	1,258	942

Standard errors clustered by MSA are in parentheses. i indicates the number of years forward for which the effect is estimated. The fixed effects, controls and instrumental variable are as in Table A4. The top panel uses single-family prices and the bottom panel prices for all homes, from Zillow. The sample period is 2009-2017. Each observation is an MSA-year. ***significant at the 1% level; **significant at the 5% level.

Table A7. Placebo panel: Housing price growth and investors' share pre-crisis

	Price growth$_{m,t}$			
Panel period	2001-2005	2001-2006	2001-2005	2001-2006
Investors' share$_{m,t-1}$	-1.465	-0.007	-0.034	-0.791
	(0.945)	(0.406)	(0.567)	(0.665)
Estimation	IV	IV	IV	IV
Instrumental variable period	2001-2005	2001-2006	2009-2013	2009-2014
Observations	1,585	1,906	1,584	1,905

Standard errors clustered by MSA are in parentheses. The specifications include location and time fixed effects and MSA-year level controls: the real housing price growth, population growth, median income growth and unemployment rate change from time $t-2$ to $t-1$. Prices are for the median house and are inflation adjusted. The instrument for the investors' share of purchases is the average share of business income over total income of the top earners in MSA m in the year 2007 multiplied by the lagged CD rate growth. In the first two columns the instruments are constructed using CD rate growth$_{m,t-1}$, so the CD rate is contemporaneous to the panel variables. In the last two columns the instruments are constructed using CD rate growth$_{m,t+7}$, so the instrument is identical to the baseline panel specification, which begins in the year 2009. Each observation is an MSA-year.

Table A8. Dynamic results controlling for labor demand shifts by industry

	Price growth$_{m,t+i}$						
	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$
Investors' share$_{m,t-1}$	0.50***	0.83***	0.69***	-0.40***	-0.74***	-0.35***	-0.71***
	(0.19)	(0.23)	(0.19)	(0.14)	(0.18)	(0.13)	(0.27)
Observations	2,756	2,440	2,135	1,834	1,529	1,224	914

Standard errors clustered by MSA are in parentheses. i indicates the number of years forward for which the effect is estimated. The specifications include location and time fixed effects and controls as in Table A4, and are estimated using our IV. Additional controls are the lagged growth rate of employment in the main industries - based on the NAICS 2 digit sector codes - within the MSAs: Health Care & Social Assistance, Retail Trade, Accommodation & Food Services, Manufacturing, Professional, Scientific and Technical Services, Administrative and Support and Waste Management, Finance and Insurance, Wholesale Trade, Other Services, and Transportation and Warehousing. Prices are for mid-tier houses and are inflation adjusted. The sample period is 2009-2017. Each observation is an MSA-year. The Kleibergen and Paap (2006) underidentification test has p-value of 0.001, and the Kleibergen and Paap Wald F statistic is 27.562 for the panel regression ($i=0$). ***significant at 1%.

Table A9. Omitted variable bias test: Oster (2019) bounds

	β ($\delta = 0$)	β^* ($\delta = 1$)	
		$R_{\max} = 1.3\tilde{R}$	$R_{\max} = 2.2\tilde{R}$
Table 3	0.0343	0.0319	0.0245
Table 7	0.0269	0.0221	0.0077
Table 8	0.0348	0.0325	0.0256
Table 9	0.0283	0.0240	0.0109

This table shows the identified intervals for the coefficient of the investors' share, based on the Oster (2019) methodology. The first omitted variable bias test is for the baseline specification (Table 3, first column), the second is for our specification with additional controls for economic drivers (Table 7, last column), the third for the specification that controls for credit denials (Table 8, first column) and the fourth for our robustness check using controls for changes in industry employment (Table 9, last column). β is the estimated coefficient in the fully controlled models. For detailed description of the methodology and the symbols check Oster (2019).

Table A10. The instrumental variable and drivers of housing markets

	Top earner business income share$_{m,07}$
Avg. median age change$_{m,00-06}$	0.019
	(0.048)
Avg. homeownership rate change$_{m,00-06}$	-0.002
	(0.055)
Median age change$_{m,07}$	-0.014
	(0.051)
Homeownership rate change$_{m,07}$	-0.024
	(0.066)
MSA-level controls	Yes
State dummies	Yes
R-squared	0.528
Observations	288

Heteroskedasticity robust standard errors are in parentheses. All variables are normalized to have zero mean and standard deviation of one. The outcome variable is our instrument for the investors' share of purchases: the average share of business income over total income of the top earners in MSA m in the year 2007. The controls are as in Table 3. Each observation is an MSA.

Table A11. Robustness: Control for individual investors

	Price growth$_{m,t+i}$						
	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$
Legal entity investors'							
share$_{m,t-1}$ (instrumented)	0.54***	0.87***	0.72***	-0.48***	-0.79***	-0.40***	-0.69**
	(0.20)	(0.24)	(0.20)	(0.15)	(0.18)	(0.14)	(0.27)
Individual investors'							
share$_{m,t-1}$ (not instrumented)	0.05**	0.06*	0.05*	-0.02	-0.06*	-0.03	-0.10
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.09)
Observations	2,842	2,525	2,207	1,891	1,575	1,258	942

Standard errors clustered by MSA are in parentheses. i indicates the number of years forward for which the effect is estimated. The legal entity investors' share is the usual definition of investors used throughout the paper. The individual investors' share is the MSA share of housing purchases in dollar value by individuals who purchase two or more properties in the same MSA within two years. The fixed effects, controls and instrumental variable for the legal entity investors' share are as in Table A4. The individual investors' share is not instrumented. The sample period is 2009-2017. Each observation is an MSA-year. ***significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table A12. Robustness: Control for foreclosures

	Bottom Tier	Mid Tier	Top Tier
	Price growth$_{m,t+1}$		
Investors' share$_{m,t-1}$	2.218***	1.120**	0.856**
	(0.803)	(0.518)	(0.380)
Number of MSAs	82	84	84
Observations	503	512	512

Standard errors clustered by MSA are in parentheses. These panel regressions are estimated at the 1-year horizon. The fixed effects, controls and instrumental variable are as in Table A4. In addition, all models control for the lagged increase in the foreclosure rate. The sample period is 2009-2017. Each observation is an MSA-year. ***significant at the 1% level; **significant at the 5% level.

Table A13. Alternative measures of investors

	Price growth$_{m,t+i}$						
	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$
Share of number of purchases$_{m,t-1}$	0.85***	1.72***	1.53***	-1.13***	-1.91***	-0.86***	-2.10**
	(0.30)	(0.48)	(0.44)	(0.39)	(0.52)	(0.32)	(1.05)
Observations	2,842	2,525	2,207	1,891	1,575	1,258	942
Share of number of units$_{m,t-1}$	0.74***	1.30***	1.15***	-0.83***	-1.35***	-0.64**	-1.12**
	(0.28)	(0.39)	(0.38)	(0.29)	(0.39)	(0.26)	(0.54)
Observations	2,842	2,525	2,207	1,891	1,575	1,258	942

Standard errors clustered by MSA are in parentheses. i indicates the number of years forward for which the effect is estimated. The investors' share of number of purchases denotes the market share of the count of purchases by investors. Each purchase counts as one purchase, independent of the type of property, that is, one single-family detached home, one apartment building, etc. The investors' share of number of units denotes the market share of the count of units purchased by investors. For example a purchase of a 10-unit apartment building counts as 10 units. The number of units is coded by ZTRAX. The online appendix describes our coding of this variable when there are missing or incomplete data from ZTRAX. The fixed effects, controls and instrumental variable are as in Table A4. The sample period is 2009-2017. Each observation is an MSA-year. ***significant at the 1% level; **significant at the 5% level.

Table A14. Land use and buildings' classification

Single-family: single family residential; townhouse; row house; mobile home; cluster home; seasonal, cabin, vacation residence; bungalow; zero lot line; patio home; manufactured, modular, prefabricated homes; garden home; planned unit development; rural residence; residential general; inferred single family residential.
Multi-family: condominium; cooperative; landominium; duplex (2 units, any combination); triplex (3 units, any combination); quadruplex (4 units, any combination); apartment building (5+ units); apartment building (100+ units); high-rise apartment; garden apartment, court apartment (5+ units); mobile home park, trailer park; dormitory, group quarters (residential); fraternity house, sorority house; apartment (generic); multifamily dwelling (generic any combination 2+); boarding house rooming house apt hotel transient lodging; residential condominium development (association assessment); residential income general (multi family).

This table shows the classification of homes into single-family and multi-family based on the ZTRAX land use standard codes.[24]

[24]We excluded from the data the following land use standard codes that do not refer to homes: "residential common area", "timeshare", "residential parking garage" and "miscellaneous improvement".

Table A15. Investors database construction

Action to clean database	Number of observations	Percentage dropped	Explanation
Database: All transactions with buyer names	226,645,766		This is the full database of deeds up to 31st December 2017, after we merged it with the buyer names based on a unique deed identifier.
Remove missing transaction dates	226,643,278	0.001%	The transaction date is the date the deed for the transfer of the property was signed.
Keep dates from 1st January 2000 to 31st December 2017	188,006,472	17.05%	The date each county begins reporting data varies. Some report from 1980, while most from the 1990s. We keep transactions from 1st January 2000.
Keep only residential properties	142,727,896	24.08%	Each property has a land use classification code. We keep the codes for single- and multi-family homes as we show in Table A14.
Remove life estate deeds	142,713,308	0.01%	This type of deed transfers the house in the future at death of the owner. Date of transfer is unknown.
Remove deeds with cancel indicator	142,608,773	0.07%	These deeds cancel previous deeds. They don't transfer ownership.
Merge 2017 MSA and MSA division codes using FIPS county codes	131,856,802	7.54%	The dropped counties are outside the Metropolitan Statistical Areas. We keep MSAs to focus on urban areas.
Merge with seller	131,687,294	0.13%	We drop incomplete records that miss the seller name.
Remove transactions with price<$10,000	85,398,628	35.15%	This is common practice with deeds data (Stroebel 2016; Bernstein, Gustafson and Lewis 2019). Most of the dropped data are from non-disclosure states.

This table describes step-by-step the cleaning of the transaction-level database. Each observation is a transaction to transfer the ownership of a property, as recorded in the official deed. The percentage dropped shows the percentage of observations that are removed from the database in each step. The cleaned database contains the dollar value of purchases by either investors or other buyers, which is used to calculate their share of purchases.