Arbeitskreis Quantitative Steuerlehre
Quantitative Research in Taxation – Discussion Papers

Marcel Fischer, Roland Füss, Daniel Ruf, Simon Stehle

Betting on Homes

arqus Discussion Paper No. 294
June 2022

www.arqus.info
ISSN 1861-8944

s:fi RESEARCH

Swiss Finance Institute Research Paper Series N°22-73

Betting on Homes

Marcel Fischer
Copenhagen Business School

Roland Füss
University of St. Gallen and Swiss Finance Institute

Daniel Ruf
Goethe University

Simon Stehle
University of Konstanz

Betting on Homes*

Marcel Fischer† Roland Füss‡ Daniel Ruf§ Simon Stehle¶

This version: June 22, 2022

*We are grateful for the many useful comments and suggestions from Brent Ambrose, Alexander Braun, Martin Brown, Andra Ghent, Jens Jackwerth, Axel Kind, Winfried Koeniger, Steffen Sebastian, Florian Weigert, seminar participants at the Universities of Konstanz, Regensburg, St. Gallen, and the Indian Institute of Management Ahmedabad (IIMA), as well as conference participants at the 2021 Annual Meeting of the German Finance Association (DGF). Marcel Fischer and Simon Stehle gratefully acknowledge financial support from the German Research Association (DFG), grant FI2141/5-1. An early version of this paper was circulated under the title: "Investor Performance in the Housing Market."

†Copenhagen Business School, Department of Finance, Solbjerg Plads 3, DK-2000 Frederiksberg, Denmark and University of Konstanz, Department of Economics, Postbox 147, D-78457 Konstanz, Germany, phone: +49 7531 882645, e-mail: marcel.fischer@uni.kn.

‡Swiss Institute of Banking and Finance (s/bf), University of St. Gallen, Unterer Graben 21, 9000 St. Gallen, Switzerland and Research Associate at the Center for Real Estate and Environmental Economics, NTNU Business School, Trondheim, Norway; Phone: +41 (0)71 224 7055, Fax: +41 (0)71 224 7088; Email: roland.fuess@unisg.ch

§House of Finance, Goethe University, Frankfurt, Campus Westend, Theodor-W.-Adorno Platz 3, 60629 Frankfurt am Main, Germany, phone: +49 (0)69 798-33645, e-mail: ruf@finance.uni-frankfurt.de.

¶University of Konstanz, Department of Economics, Postbox 147, D-78457 Konstanz, Germany, phone: +49 7531 88 3620, e-mail: simon.stehle@uni.kn.

Abstract

Betting on Homes

In this paper, we analyze the differences in annualized capital gains across heterogeneous investor groups in the US housing market, namely owner-occupiers, private investors, as well as short- and long-term institutional investors. Our empirical results link the performance differences to heterogeneity in risk-taking. In particular, investor-specific exposure to lagged local return dispersion explains persistent performance differences of investors within a given market. Short-term institutional investors outperform others by exploiting the upside potential of local return dispersion. By contrast, neither macroeconomic fundamentals nor local factors, such as momentum and downside risk, can explain the observed performance disparities.
JEL Classification Codes: R30, R32

Key Words: Capital gains; housing market; institutional investors; local markets; risk exposure; upside potential.

1 Introduction

Residential real estate differs from traditional asset classes, such as stocks or bonds, in many ways. For instance, while it is notoriously difficult to predict future stock prices, housing markets are generally subject to strong and positive momentum (Case and Shiller, 1989). Additionally, homes are indivisible and comparably expensive, making housing risk difficult to diversify. Simultaneously, the housing market is occupied by a wide range of buyers with contrasting strategies and targets, varying, for example, in investment horizon, risk-taking, and whether a home is bought primarily as a consumption good or as a financial asset. Consequently, buyers' and sellers' ties to local market returns and housing risk are likely to be highly divergent, suggesting that heterogeneous exposure to risk is crucial in understanding investment behavior and expected returns in these markets.

In this paper, we analyze the role of local uncertainty regarding capital gains, in order to understand performance differences between different investor groups in the residential housing market. We study the heterogeneous exposure of investors' trades to such uncertainty, measured by the local dispersion of realized capital gains in a given period relative to nationwide dispersion. Building on our first result that the heterogeneous exposure to local return dispersion helps to explain the otherwise systematically different investor performances, we offer new insights into housing market speculation. Specifically, we show that institutional investors conducting short-term trading dynamically select markets with high and increasing return dispersion and profit particularly from the upside potential of such markets when selling. Hence, our findings suggest a specific strategy of such investors, based on betting on the upside potential, which is rewarded with higher expected capital gains.

Our analyses are based on four prevailing categories of market participants, that are likely to be heterogeneous in their main objectives as well as their underlying strategies. Our first category consists of owner-occupiers (OO), whose main interest should lie in housing consumption rather than earning financial yields. As consumption implies selection based on preferred location and taste, the compensation for engaging in riskier trades should be of secondary importance to this group. By contrast, we expect primarily financially motivated investors to seek compensating (expected) returns through increased risk-taking. These investors can be classified into three categories. First, private investors (PI) might buy second homes as future savings for retirement, but were also active during the last housing boom (e.g., Bayer et al., 2020, 2021), suggesting high exposure to momentum and risk factors. Second, short-term institutional investors (SI) are likely to speculate on high appreciation rates and should thus be highly exposed to local momentum and risk. Third, long-term institutional investors (LI) should be more likely to trade on predictable rental income, e.g., to conduct asset-liability management, which indicates a relatively lower exposure to momentum and housing market risk, respectively.

We analyze more than 21 million repeat sales of US residential real estate, which we relate uniquely to each of the investor types defined above. As our first result, we document persistent and sizable performance differences among the four groups of market participants. On average, SIs systematically outperform all other investor types not only cross-sectionally, but also within local markets over the sample period. Our finding is robust against a set of control variables and several fixed effects specifications. The disparities in realized capital gains can be explained neither by unobserved cross-sectional heterogeneity across markets, individual holding periods, construction activity, nor investment timing at the aggregate or local level. Our results therefore hint at investor-specific factor exposure and related strategies, which lead to persistent performance differences, even within the same market and the same state of the housing market cycle.

Our next step is thus to show that local return dispersion prior to selling can explain the observed capital gain disparities between the investor groups. Our performance analyses are based on the generalized portfolio sorts model proposed by Hoechle et al. (2020), which allows us to simultaneously test the relative performance and risk factor exposure of investor groups in the same local market. Our regressions show that the relative alphas become insignificant after including lagged local return dispersion. Furthermore, we observe a negative exposure to local return dispersion at the time of sale for all investor groups, except for SIs, who seem to be rewarded with higher realized capital gains. Motivated by the revealed heterogeneity in investor exposure, we then present the upside variation of observed capital gains prior to selling, as a predictor of SI outperformance. Our results suggest that SIs trade strategically on the upside potential in local markets to outperform other investor groups.

Overall, our findings remain intact when we control for local return dispersion prior to the purchase decision, as well as momentum at purchase and sale date, further improving comparability between investors conditional on their entry and exit at different states of the housing market cycle. The results are also robust to instrumenting return dispersion similarly to Han (2013), alternative definitions of "short-term activity", and different subsamples. By contrast, the relative performance differences of investors remain large and significant when testing alternative channels, such as local or nationwide momentum, compensation for downside risk, liquidity, or macroeconomic fundamentals in a local market.

Our findings suggest that local return dispersion prior to a sale can be exploited to explain the SI outperformance. However, this raises the question of whether SIs are compensated with higher capital gains for betting on the upside potential in local markets as a persistent strategy. If this is the case, we expect them to be capable of identifying such markets already on making their purchase decision. We therefore continue our analysis by studying potential determinants of market selection across investor types. Specifically, we show that the SI outperformance can also be predicted at the beginning of the investment decision by exploiting lagged local return dispersion as well as contemporaneous dispersion growth. This finding is consistent with our intuition. Return dispersion growth observed at the time of market entry serves as an additional predictor of potential capital gains, which are primarily targeted over a short-term holding period.

Our final step is to employ an instrumental variable approach to identify a possible strategy of SIs that connects their persistent outperformance to the local upside potential. In particular, we hypothesize that SIs systematically buy relatively cheaper properties from financially distressed sellers, while simultaneously betting on the future upside potential of the respective market. To instrument distressed markets, we make use of the 2008/2009 LIBOR-Treasury shock used by Gupta (2019). Specifically, we exploit the fact that markets with a particularly high 2007 share of LIBOR (compared to Treasury) linked ARMs were relatively more exposed to the unexpected increase in the LIBOR-Treasury spread due to mortgage payment shocks. Supporting our hypothesis, we find that SI activity is higher in such markets after the 2008/2009 shock, and that the corresponding instrumented SI activity positively and significantly predicts an increase in upside potential realized upon sale, while downside risk, in contrast, remains statistically unrelated to our instrument. Our results thus suggest that SIs manage to outperform other investors by buying cheaper, distressed homes and betting on subsequent local upside potential.

The key quantity of our analysis consists of annualized capital gains in excess to the risk-free rate. For several reasons, we focus on this measure rather than rental income. First, capital gains are of interest to all four types of investors. OOs should also have an interest in homes with high appreciation rates, as they benefit from increased consumption (Campbell and Cocco, 2007), entrepreneurship (Corradin and Popov, 2015), higher local economic growth (Loutskina and Strahan, 2015), as well as from a hedge against future housing consumption risk (Han, 2013). Second, capital gains can be observed directly for all trades. In contrast, rents must either be imputed for the largest share of homeowners, as OOs are by far the largest group, or used at the aggregate level, in which case they can be addressed with fixed effects. Additionally, in efficient markets, expected rental income should be reflected in observed trading prices. Thus, the comparatively stable rental income is further addressed with our repeat sales approach and additionally captured by local fixed effects.

Our work provides a unifying framework to analyze and compare performance and risk exposure of different types of investors that have been investigated in prior literature.

For instance, similar to our category of PIs, second-home buyers have been analyzed by Chinco and Mayer (2016) and Cvijanovic and Spaenjers (2021), focusing on out-of-town and foreign buyers, respectively. Bayer et al. (2021) document investor contagion from experienced investors to new, inexperienced ones, who then performed worse than their professional counterparts. Speculators, typically engaging in short-term activity, have been extensively analyzed in the literature (e.g., Fu and Qian, 2014; Fu et al., 2016; Bayer et al., 2020). For instance, short-term speculators drive up trading volume, as documented by DeFusco et al. (2022). Institutional investors further speed up the recovery of housing markets after the recent bust, as shown by Lambie-Hanson et al. (2022). Mills et al. (2019) analyze single-family home buyers, i.e., long-term investors, who should be interested in rental income and operate on a larger scale, so as to securitize the respective payment streams.

We contribute to the literature on investors in the housing market by showing that a sizable share of the systematic differences in capital gains across prevailing investor groups can be linked to heterogeneous risk exposure. Our analyses not only focus on (selected) MSAs, but on the entire US mainland, over a period of more than two decades, allowing us to identify local return dispersion, and specifically betting on the upside potential, as a novel explanation of performance differences, which are persistent over local boom and bust cycles. Simultaneously, we show that trading on momentum can be beneficial in general, but explains little of the observed disparities across investor groups. Thus, we can shed light on the effects of location choice, timing on aggregate as well as local markets, associated risks and returns. Our findings thus not only have implications for the most densely populated areas, but also for more rural regions. Moreover, our results are not only relevant for a particular investor group, but for a wide range of homeowners in the US.

By highlighting the importance of local return dispersion for the understanding of investor outperformance relative to OOs, we also contribute to a growing strand of literature that analyzes the relationship between risk and return in the housing market. Han (2010) shows that incentives to buy a home as a hedge against future housing risk can influence the demand for homes in local markets, thus revealing a heterogeneous impact of risk on local markets. These hedging incentives further affect the risk-return relationship across metropolitan statistical areas (MSAs). As documented by Han (2013), hedging incentives against future housing risk can even lead to a negative risk-return relationship in some areas. Similarly, Peng and Thibodeau (2017) find that, on average, idiosyncratic risk is not compensated for by higher appreciation rates. We contribute to this strand of literature by relating local return dispersion to the activity of various classes of investors as well as their underlying performance in capital gains. Our results document that even within the same markets, investors' risk exposure is heterogeneous, which can help explain the ambiguous relationship between risk and return in housing markets.

The remainder of this paper is structured as follows. In Section 2, we describe our data, the cleansing process, and introduce the key variables. Our methodology and results are presented in Section 3. Section 4 provides robustness checks and Section 5 concludes.

2 Data

In this section, we present our data. Section 2.1 briefly discusses the data cleansing process. In Section 2.2, we explain the identification of investor types. Section 2.3 describes our key performance variable, annualized excess capital gains from repeated homes transactions, as well as our measure for local return dispersion. Section 2.4 introduces the control variables. Finally, we present summary statistics in Section 2.5.

2.1 Housing transactions

We obtain a representative dataset of residential housing transactions from data vendor *CoreLogic*, whose databases cover more than 99% of US properties.[1] Our data cleansing process starts with 99,757,949 arms-length transactions with respect to apartments, condominiums, duplexes, and single-family homes with positive sales prices and complete sales dates from 1995 to 2017.[2] To remove duplicate transactions, we follow DeFusco et al. (2022). First, if there is more than one transaction with coinciding buyer and seller names and identical sales prices, we only retain the transaction with the earliest recording date. Second, if the same property is traded many times on a given day, we keep only one of the given transactions (85,966,973 remain). To account for price outliers likely to be associated with data errors, we follow Bollerslev et al. (2016) and remove all transactions outside an interval with fixed nominal bounds of 5,000 USD and 100,000,000 USD (85,641,056 remain).

Of the remaining transactions, using the unique property identifier provided by *CoreLogic*, we identify 33,280,346 repeat sales. As suggested by Bollerslev et al. (2016), we exclude extreme observations by first dismissing capital gains of less than -50% or more than 100% per year, and second, removing repeat sales with a holding period of less than 180 days (29,748,815 remain). Observations for which an investor identity cannot be assigned, e.g., due to a missing address of a private homeowner, are also dismissed from the sample (26,217,238 remain). We then dismiss annualized capital gains in the first and last percentile of our

[1]As stated by the data vendor on https://www.corelogic.com/find/property-data-solutions/. Last retrieved on June 20, 2022.

[2]Table A1 in the Online Appendix (currently at the end of this manuscript) summarizes all steps of the data cleansing process.

sample (25,692,895 remain). At the final stage of our cleansing process, we remove real-estate-owned (REO) related repeat sales. First, as these foreclosures typically fail as auction sales and are therefore subject to large price discounts (e.g., Campbell et al., 2011; Anenberg and Kung, 2014; Gerardi et al., 2015), repeat sales not flagged as REO-transactions, but with similar or more extreme capital losses, can more plausibly be identified as irregular. Second, removing them before matching repeated sales would lead to "false" transaction pairs.[3] The final sample contains 21,178,869 observations.

2.2 Defining investor types

We define four mutually exclusive investor types that differ in their investment strategies. Our first group encompasses owner-occupied households (OOs), i.e., buyers who live in the home underlying the transaction. Second, we define a private investor (PI) category that consists of second-home buyers (e.g., Chinco and Mayer, 2016; Cvijanovic and Spaenjers, 2021; Gao et al., 2020). Third, short-term operating professionals (SIs) with their main interest in capital gains, and fourth, long-term professionals (LIs), such as large-scale single-family home buyers who focus on steady rental income (Mills et al., 2019). This approach is in line with the recent literature associating speculation with short holding periods (e.g., Bayer et al., 2020, 2021).

Technically, we assign repeat sales to one of the four investor types as follows. We first distinguish between corporate and private identities, linked to the purchase of repeated transactions, using an identifier provided in the *CoreLogic* dataset. To divide private buyers into OOs and PIs, we compare homeowners' mailing addresses with the corresponding property addresses (DeFusco et al., 2022). As the primary indicator of owner-occupancy status, we check whether house numbers of both addresses match. If at least one of the two house numbers is missing, we compare street names instead. If the respective criterion indicates a match, owner-occupancy status is assigned, and PI identity otherwise. If insufficient information for identification is available, we omit the observation from the sample. Corporate identities are referred to as institutional investors, which are further divided into either short-term (SI) or long-term institutional investors (LI), depending on the respective holding period. We follow DeFusco et al. (2022) and define a holding period of less than or equal to three years as short-term, and assign long-term status otherwise.[4]

[3]For instance, consider a property that is traded three times, with the second transaction being a foreclosure. An early omission of the foreclosure would lead to a misleading match between the first and third sale, and thus implies a false capital gain as well as a falsely assigned investor identity.

[4]In Section 4 on robustness tests, we show that our base case results are robust to alternative cut-off values at two or four years, respectively.

2.3 Construction of key variables

In this paper, we study systematic differences in capital gains across investor types. However, realized returns are difficult to compare cross-sectionally, as they occur over different holding periods and originate between different points in time. Thus, we make two adjustments to realized returns similar to Fischer et al. (2021). To account for varying holding periods, we first annualize each capital growth rate. Additionally, we correct for time-varying opportunity costs by subtracting the annualized return of an investment at the risk-free rate over the same period. Accordingly, we derive annualized excess capital gains, $r_{it_1t_2}$, of home i, bought at t_1 and sold at t_2 as

$$r_{it_1t_2} = \left(\frac{P_{it_2}}{P_{it_1}}\right)^{\frac{1}{h(t_1,t_2)}} - \left(R^f_{t_1t_2}\right)^{\frac{1}{h(t_1,t_2)}}, \tag{1}$$

in which P_{it} is the nominal sales price at time t, $h(t_1, t_2)$ is the holding period in years, and $R^f_{t_1,t_2}$ is the gross risk-free rate from t_1 to t_2.

We then derive our key variable for local return dispersion as a proxy for uncertainty about realized capital gains in a specific market. To achieve a reasonable tradeoff between estimation precision and cross-sectional as well as temporal variation, our measure is derived at the county-quarter level. Otherwise, the use of more fine-grained frequencies or areas might lead to the loss of more rural areas in the sample, which would bias our results towards large MSAs.

Local return dispersion is derived as the standard deviation of all annualized capital gains that are realized in a given period in a given location. To account for an aggregated common market trend in the overall return risk, we standardize the local measure by the nationwide standard deviation. Our return dispersion measure thus allows for a cross-sectional comparison, indicating whether the local dispersion of capital gains is relatively high or low, compared to the nationwide variation.[5] Formally, our measure for local return dispersion, D_{ct}, is derived as

$$D_{ct} = \frac{\sigma_{ct}}{\sigma_t}, \tag{2}$$

in which σ_{ct} and σ_t are the standard deviations of annualized returns realized in county c as well as the nationwide level at time t.

[5] Because we compute return dispersion as a proxy for uncertainty about cross-sectional capital gains in local markets, our procedure does not rely on prior index aggregation, but directly exploits the underlying data. Also, our approach differs, for instance, from Han (2013), whose measure depends on implicit GARCH model assumptions on the autoregressive structure, in order to study the time series implications of local market uncertainty.

2.4 Control variables

We investigate channels other than local return dispersion by making use of additional data sources and variables derived from our dataset. To capture the exposure to risk factors and market conditions at the purchase and sale date, we match all variables to both transaction dates of each observation, respectively.

Based on our housing dataset, we calculate additional explanatory variables at the county-quarter level. First, we want to control for house price growth as a momentum-based trading channel and a potential explanation of the performance differences between investors. To capture local price dynamics, we compute county-level indices using the Case and Shiller (1989) methodology, exploiting the repeat sales structure given in our data.[6] We measure turnover as the sum of all individual transactions in a given period, to proxy local market liquidity. Local credit conditions are captured with the average primary loan-to-value ratio (LTV), derived as the average mortgage amount divided by the average sales price, excluding individual ratios larger than one in the mean calculation, similar to Fuster and Vickery (2015). Intuitively, this measure allows us to control for potential exposure to the debt level in the local housing market.

Local macroeconomic fundamentals are available at the county-year level. We use information on median income, the unemployment rate, population, and land area to calculate population densities. Population and land area (measured in square miles) are obtained from the US Census Bureau. Median income is provided by the Small Area Income and Poverty Estimates Release of the US Census, obtained from the FRED database of the Federal Reserve Bank of St. Louis. The unemployment rate is provided by the US Bureau of Labor Statistics.

We also use monthly time series of established macro-financial risk factors. To investigate investor exposure to the aggregate housing market, we refer to monthly data of the S&P Case-Shiller US National Home Price Index (not seasonally adjusted) and the average 30-year mortgage interest rate, both obtained from FRED. We further attain stock market returns, as well as data on the risk-free rate from Kenneth French's webpage.[7]

Exploiting the abundance of our data, we also control for transaction-specific characteristics. Assuming a constant quality over time, the potential impact of time-invariant

[6] When estimating price indices, we face the problem of missing observations during in-between periods for some counties. To make use of the maximum amount of observations possible, we solve this issue by omitting the respective intervals in the estimation process and assigning missing values to these periods in the resulting time-series. To further increase the precision of our estimates, we use data on transactions ranging back to 1980 when possible, as well as transactions with unknown owner identity, such that indices are estimated before dismissing these observations in the cleansing process.

[7] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html, last retrieved on June 20, 2022.

hedonic characteristics on capital gains, is removed by the repeat sales structure of our data. However, when investigating capital gains among investor types, larger renovations might confound the true performance, because observed gains might be driven upwards from additional investments, rather than reflecting the general price appreciation. To mitigate potential concerns in this respect, we therefore construct a dummy that indicates the existence of construction-related documents (e.g., a construction mortgage) associated with the property purchase. Similarly, to rule out the performance differences being driven by real estate developers selling newly built homes to the market (Nathanson and Zwick, 2018), we generate an additional dummy indicating that our data provider classifies the initial purchase as new construction. We further control for unobserved heterogeneity related to the holding period. For instance, shorter holding periods are documented as generally associated with speculative activity, thus possibly implying higher returns (Bayer et al., 2020). To account for a potentially non-linear relationship, and to improve the comparability between investor types, we generate dummies for sample deciles of different holding periods that we include in our regressions.

2.5 Descriptive statistics

Table 1 provides summary statistics of the sample. Panel A shows that the sample average of annualized excess capital gains is 4.2%, reflecting the general increase in market prices over the sample period between 1995 and 2017. This finding is further reflected in the comparison of nominal purchase and sale prices, with an increase in average values from about 250,000 USD to 310,000 USD. A comparison with the median indicates the typical right-skewness of the capital gains distribution. Almost one percent of purchases is associated with larger observed construction activity. The number of observations of our local expected return (momentum) and dispersion measures is reduced, compared to our overall sample, and for two reasons. First, we account for extreme values by trimming observations out of the 0.5 to 99.5 percent interval. Second, as we use lagged versions of these variables in our analyses to cope with potential simultaneity issues, we report summary statistics of the time-shifted variables for the sake of consistency.

Panel B documents average annualized excess capital gains, as defined in Equation (1), broken down by investor group. With an average excess capital gain of about 3.5%, OOs perform worst among the groups. Yet, the corresponding standard deviation indicates that, as a group, OOs face the lowest variation in excess capital gains among the investors. The risk-return structure holds for all subgroups: A higher group average is associated with a higher standard deviation. The highest gains (15.8%), but also the highest dispersion (24.2%) is realized by SIs. This finding might already indicate that risk-taking, reflected in the varying return dispersion, might contribute to the different performance levels across investors. However, as suggested by the Sharpe ratios, SIs also outperform other market participants in terms of higher realized excess capital gains per level of risk taken by investors.[8]

In Figure 1, we demonstrate that the investor-specific differences in annualized excess capital gains not only hold cross-sectionally, but are also persistent over time. SIs performed best since 1998. By contrast, OOs persistently earned the lowest capital gains since 1996. Furthermore, the graph suggests a common trend in average capital gains realized by investor types during the housing boom, raising the question of whether investors generally followed similar strategies. However, when comparing the performance before and after the housing market bust in 2007, capital gains earned by SIs started to increase again, reaching and even surpassing their pre-crisis level in 2014. The other investor groups obtained negative capital gains on average, reaching a lower common trend after the burst.[9] Having presented descriptive evidence from our data, we provide a more formal analysis on the documented performance differences in the following section.

3 Empirical results

In this section, we present our results. In Section 3.1, we discuss the methodology for analyzing relative performance and investor-specific exposure to local factors. In Section 3.2, we use simple regressions to support the descriptive evidence on the performance differences across investor groups. Section 3.3 then documents that lagged local return dispersion explains disparities across investors' capital gains. In Section 3.4, we further disentangle return dispersion and show that the upside potential predicts the outperformance of SIs. In Section 3.5, we then study the role of local return dispersion in the purchase decision for market participants, and identify a trading strategy of SIs using an instrumental variable approach in Section 3.6.

[8] In Online Appendix A (currently at the end of this manuscript), we provide further evidence of the observed performance differences. For instance, Figure A1 shows that (short-term) institutional investors consistently outperform the other investor groups, compared within the same holding-period deciles. In Figure A2, we document heterogeneity in investor shares of debt-financing, construction activities, or purchases of newly built homes. However, Figure A3 suggests that performance differences are persistent, even when conditioning on these variables.

[9] Figure A4 in Online Appendix A (currently at the end of this manuscript) confirms the performance persistence of SIs for different housing market sectors. By contrast, all other investor groups underperformed after the bubble burst in 2007. As suggested by Figure A5, SIs also outperform other investors, when improving their comparability by using only capital gains realized over a maximum holding period of three years.

Table 1
Summary statistics

Panel A: Full sample

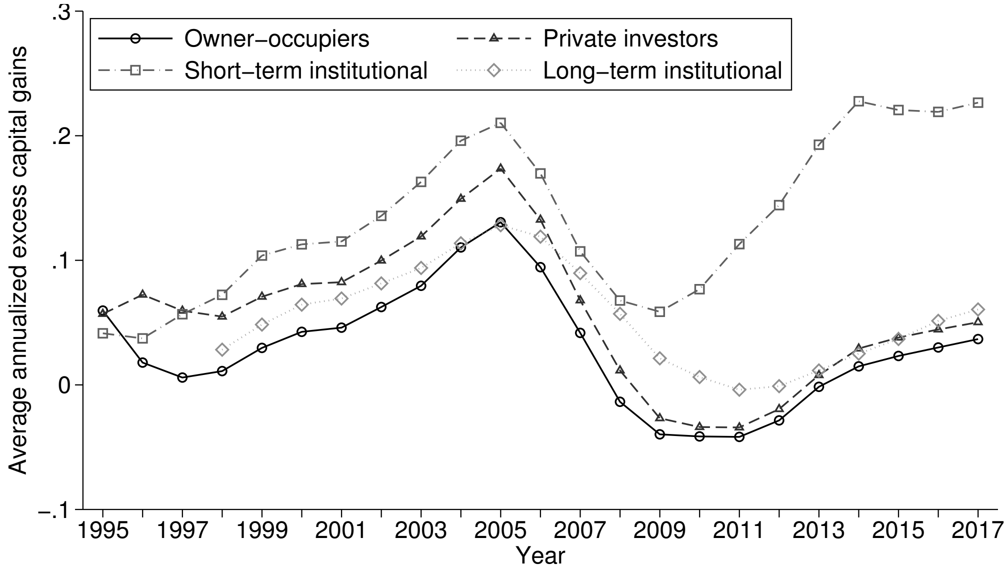
	Observations	Mean	Std.	Median
Ann. excess capital gain	21,178,869	0.042	0.121	0.019
Purchase price ($)	21,178,869	253,182	486,727	179,037
Sale price ($)	21,178,869	313,509	677,930	220,000
Construction dummy	21,178,869	0.008	0.088	0
New construction dummy	21,178,869	0.158	0.365	0
Holding period (years)	21,178,869	5.750	4.005	4.712
County unemployment (percent)	21,170,934	5.341	2.144	4.900
County income ($)	21,156,784	49,813	12,767	47,215
County pop. density (pop. per sq. mi)	21,178,517	1,679	4,890	685.966
County primary LTV (percent)	21,178,545	0.638	0.168	0.682
County turnover (#)	21,178,869	4,021	5,169	2,019
County momentum (percent)	20,825,036	0.017	0.039	0.019
County return dispersion	19,760,762	0.950	0.264	0.931

Panel B: Annualized excess capital gains by investor

	Observations (%)	Mean	Std.	Median	Sharpe
Owner-occupiers (OO)	15,660,837 (73.9)	0.035	0.105	0.017	0.333
Private investors (PI)	4,590,293 (21.7)	0.055	0.144	0.023	0.382
Short-term institutional investors (SI)	435,175 (2.1)	0.158	0.242	0.120	0.653
Long-term institutional investors (LI)	492,564 (2.3)	0.052	0.139	0.024	0.370

Panel A of this table shows summary statistics for the full sample. “Ann. excess capital gain” is the annualized capital gain of a repeat sale minus the return of the risk-free rate over the respective holding period. “Purchase (sale) price” is the corresponding nominal transaction price of the first (second) sale in USD. “Construction” is a dummy that indicates whether the purchase is associated with construction-related documents, such as a construction mortgage, excluding new constructions. “New construction” is a dummy that indicates the purchase of a newly constructed home. “Holding period” is that between purchase and sale date, given in years. “County” variables indicate local market conditions related to each transactions’ purchase period. “County unemployment” is the county-year unemployment rate in percentage points, “County income” is the county-year level median income,“County pop. density” is the population density on the county-year, defined as the number of inhabitants per square-mile of land, “County primary LTV” is the county-quarter level average of the mortgage amount, divided by the average purchase price, and “County turnover” is the sum of all transactions in a given quarter in a given county, including observations with unknown identity. “County momentum” is the one-quarter lag of our estimated county-level index returns. “County return dispersion” is the one-quarter lag of the local annualized return dispersion, relative to nationwide dispersion. Panel B shows summary statistics of annualized excess capital gains and the Sharpe ratio separated by investor type. “Owner-occupiers (OO)” are private individuals who live in the underlying home. “Private investors (PI)” are private individuals who do not live in the home purchased. “SI” indicates that the property is purchased by a short-term institutional investor with a holding period of three years or less, “LI” are long-term institutional investors with a holding period of more than three years. The numbers in parentheses next to total amount of observations per investor group, indicate the relative frequency of each group in the sample.

Figure 1
Capital gains in the US residential housing market



This figure shows the evolution of average annualized excess capital gains realized at a given year by investor groups between 1995 and 2017. Investor identities are uniquely assigned to each repeated sale: owner-occupiers are defined as households which live in the traded home, private investors are households which do not live in the traded home, short-term institutional investors (SIs) are companies which sell within a holding period of three or less years, long-term institutional investors (LIs) sell after a holding period of more than three years.

3.1 Methodology

In this paper, we study local return dispersion as a predictor of the observed performance differences between investor types. As each group should follow a distinct investment strategy, the relationship between realized capital gains and lagged local return heterogeneity might be crucial in understanding their return disparities. We therefore evaluate the investor performance along two dimensions. First, we estimate the group-specific exposure to local return dispersion to infer whether the risk-return relationship is indeed heterogeneous. Second, we test whether this heterogeneity can explain the observed differences in capital gains.

To answer both questions simultaneously, we proceed similar to the generalized portfolio sorts approach proposed by Hoechle et al. (2020). First, the methodology has the advantage that relative performance alphas and factor exposures can be estimated in a single panel regression when explaining realized capital gains. Second, in contrast to the traditional portfolio sorts approach (Black et al., 1972), this specification allows for the inclusion of control variables and fixed effects, to account for unobserved heterogeneity across different local housing markets. Consequently, we explain the average annualized excess capital gain $r_{ict_1t_2}$, realized between purchase date t_1 and sale date t_2 by an investor of group k for property i that is located in county c, with regressions of the form

$$r_{ict_1t_2} = (\boldsymbol{\alpha_i} \otimes \boldsymbol{X_{ct-1}})\, \boldsymbol{\theta} + \nu_c + Z_i\beta + \epsilon_{ict_1t_2}. \tag{3}$$

Vector $\boldsymbol{\alpha_i} = \begin{bmatrix} 1 & \alpha_i^{(PI)} & \alpha_i^{(SI)} & \alpha_i^{(LI)} \end{bmatrix}$ contains dummy variables of the investor type identified for repeat sale i, $\boldsymbol{X_{ct}} = [1 \quad \boldsymbol{Q_{ct_1}} \quad \boldsymbol{Q_{ct_2}}]$ including a set of regressors, with $\boldsymbol{Q_{ct_1}}$ containing variables related to the purchase date of property i, and $\boldsymbol{Q_{ct_2}}$ to the respective sale date. In our baseline specifications, we use local return dispersion, D_{ct}, as a factor. The explanatory variables are included with a one-period lag to avoid simultaneity issues and a potential look-ahead bias that might arise, as both regressors are constructed from realized capital gains.[10] Also, the lagged specification allows us to test local return dispersion as a predictor of the investor-specific performance differences.

Vector $\boldsymbol{\theta}$ contains the regression coefficients. The estimates for the interactions of $\alpha_i^{(PI)}$, $\alpha_i^{(SI)}$, and $\alpha_i^{(LI)}$ with the "intercept" in $\boldsymbol{X_{ct}}$ measure the factor-adjusted performance of each investor type, relative to OOs as the reference group ("relative alpha"). An insignificant coefficient estimate associated with the corresponding investor type therefore indicates that the outperformance relative to the reference group can be explained by the set of (local) factors included in the regression. Parameter ν_c captures unobserved heterogeneity across local house price developments at the county level, potentially arising from time-invariant factors such as land-supply restrictions (Saiz, 2010). Furthermore, included county-level fixed effects ensure the comparability of the performances between investor types with property investments in the same local market. Hence, we test whether local return dispersion as a time series predictor can persistently explain within-county performance differences between investors. Matrix Z_i includes additional control variables associated with transaction i, such as dummies for construction, newly built homes, and for holding-period deciles, to further improve the resemblance between investor types and observed investment horizons.

3.2 Systematic differences in capital gains

In a first step, we use simple panel regressions to show that the observed performance differences are persistent both cross-sectionally and over time. Table 2 compares the average annualized capital gains of all investor groups relative to OOs as a benchmark conditional on a set of control variables. In column (1), we regress annualized excess capital gains on

[10] In Section 4, we proceed similarly to Han (2013), and instrument contemporaneous return risk with up to two of its lags. Our results are robust to these alternative specifications.

dummy variables indicating the respective investor type. As for all performance analyses in the paper, we use OOs as the base category. On average, SIs outperform OOs by more than 12 percentage points per year, while PIs and LIs perform 2 and 1.6 percentage points better, respectively.

A large portion of the documented outperformance might stem from short-term value-creation strategies, such as renovations. To control for these effects, in column (2), we add dummy variables for observed construction activities, investments in newly built homes, or decile bins of different holding periods. We find a sizable reduction in capital gains realized by SIs, but also for PIs, relative to OOs, which might be explained by short-term value creation, such as house flipping strategies (e.g., Bayer et al., 2020), or real estate developers selling newly built properties in the market (e.g., Nathanson and Zwick, 2018; Gao et al., 2020). By controlling for the holding period, we also filter out unobserved variation in capital gains, driven by short-term trading activity. Furthermore, we show that all investor groups consistently outperform OOs when compared in the same holding period bin, with highest annualized excess capital gains realized by SIs. Hence, our findings suggest that the outperformance is not merely mechanically driven by our definition of holding periods.

Columns (3) and (4) additionally include county-level fixed effects and dummies to control for the investment timing of purchase and sale at the annual and quarterly levels, respectively. This specification captures the impact of aggregated price dynamics on investor performance when comparing their timing decisions within a county. The disparities in observed capital gains are positive and statistically significant, not only across counties, but also within local markets.

As indicated in column (5), interacting location with both purchase and sale quarters to county-quarter fixed effects can explain up to 46% of the variation in annualized excess capital gains. However, despite this sizable explanatory power when controlling for unobservable within-county factors, the return differences across investors persist. The pattern even remains after the inclusion of zip-code-quarter fixed effects in column (6). Overall, the results suggest that the outperformance of investor groups relative to OOs is largely driven by their investor-specific factor exposure within local markets. We therefore continue our analysis by investigating different potential channels using the more advanced methodology outlined in the previous section.

3.3 Can return dispersion explain the performance differences?

The next step is to consider whether the performance differences can be explained by local return dispersion. We additionally test a set of alternative channels to rule them out as

Table 2
Average differences in capital gains

	(1)	(2)	(3)	(4)	(5)	(6)
PI	0.020***	0.014***	0.018***	0.019***	0.019***	0.016***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
SI	0.123***	0.048***	0.060***	0.060***	0.067***	0.055***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
LI	0.016***	0.030***	0.032***	0.032***	0.029***	0.027***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Construction	-	X	X	X	X	X
New constr.	-	X	X	X	X	X
Holding pd.	-	X	X	X	X	X
Local FE	-	-	county	county	-	-
Purchase FE	-	-	year	quarter	-	-
Sale FE	-	-	year	quarter	-	-
L×T FE	-	-	-	-	cty-qtr	zip-qtr
Adj. R-sq.	0.024	0.135	0.324	0.328	0.467	0.519
Observations	21,178,869	21,178,869	21,178,869	21,178,869	21,178,869	21,087,407

This table shows regression results with annualized excess capital gains as the dependent variable. The displayed coefficients are estimates for group-specific dummies. The reference category are owner-occupiers (OO), i.e., private individuals who live in the purchased home during the holding period. "PI", "SI", "LI" are dummy coefficients indicating that the corresponding trade was made by a private investor, short-term institutional investor or long-term institutional investor, respectively. Private investors are defined as individuals who do not live in the underlying homes. A transaction is classified as "short-term" if the corresponding holding period is three years or less, and "long-term" otherwise. "Construction" indicates that the purchase transaction has associated documents indicating construction activity, excluding new constructions. "New construction" indicates that the purchase is identified a newly built home. "Holding period" stands for dummies based on holding-period sample deciles. "Local FE" indicates local fixed-effect dummies on the level given in the respective column. "Purchase FE" and "Sale FE" indicate that fixed-effect dummies corresponding to the respective purchase or sale date on yearly or quarterly level are included in the model, respectively. "L×T" indicate that interactions of local and time fixed-effect dummies are used, respectively. Here, "cty-qtr" denotes county-quarter, and "zip-qtr" denotes zip-code quarter fixed effects, respectively. For all local-time fixed effects, interactions of both purchase and sale quarter are used. Standard errors are clustered over counties. *,**, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively.

potential drivers. In Table 3, we first investigate the relationship between local return dispersion and investor-specific performance differences. We regress annualized excess capital gains on the investor type dummies, and the corresponding interaction terms with one-quarter lagged local return dispersion. Panel A shows that the relative investor alphas within a county can be explained by the exposure to local return dispersion at the time of a sale. For ease of comparability, column (1) displays the performance differences conditional on a set of controls. Again, we compare investors within the same holding-period deciles. Column (2) suggests that return dispersion prior to the time of purchase does not predict the alphas between investor types. By contrast, as shown in column (3), the exposure to local return dispersion prior to the sale can fully explain the outperformance of investors relative to OOs. Controlling for local return heterogeneity prior to the respective investor's market entry in column (4) leaves a borderline significant outperformance of LIs. While this outperformance appears to be unstable regarding our robustness checks, it may well indicate that the long-term strategy pays off, due to lower risk in capital gains heterogeneity relative to their performance.

In Panel B, we illustrate the estimated investor-specific exposure to local return dispersion in the purchase and sale quarter, respectively. The coefficients correspond to the model specification presented in column (4) of Panel A. The purchase loadings are positive and similar for OOs, PIs, and LIs. The coefficient signs of the factor exposure suggest a general compensation for risk-taking when purchasing the home. This finding is consistent with Han (2013), who relates increasing price uncertainty of housing investments to higher compensating returns via discounts in the purchase price. Similarly, the sale date exposure is negative, suggesting that selling under higher local market heterogeneity is generally associated with a discount. This corresponds to the compensation required by new home buyers at the purchase date, for increasing uncertainty about capital gains when engaging in markets with higher local return dispersion. The negative coefficient of sale-date risk exposure of OOs is higher than for PIs and LIs, suggesting a larger risk discount for individuals with a primary consumption motive. The exposure of SIs is remarkably different to the other groups of investors, with a large, positive exposure at the sale date. This might indicate that the trading strategy of SIs pays off most in markets with higher capital gains heterogeneity, potentially reflecting speculative activity.

3.4 Does the upside potential predict the SI outperformance?

In this section, we further investigate the channel through which local return dispersion explains the performance differences. Specifically, we study whether the heterogeneous ex-

Table 3
Investor performance and local return dispersion

Panel A: Relative investor alphas under local return dispersion exposure

	(1)	(2)	(3)	(4)
PI	0.013***	0.011**	-0.001	0.001
	(0.001)	(0.004)	(0.006)	(0.005)
SI	0.049***	0.049***	-0.017	-0.011
	(0.003)	(0.008)	(0.016)	(0.015)
LI	0.031***	0.023***	0.011	0.014*
	(0.001)	(0.005)	(0.006)	(0.006)
Local return dispersion	-	purchase	sale	purchase/sale
Construction	X	X	X	X
New construction	X	X	X	X
Holding period	X	X	X	X
County FE	X	X	X	X
Adj. R-sq.	0.161	0.170	0.166	0.175
Observations	21,178,869	19,760,762	20,931,159	19,626,314

Panel B: Local return dispersion exposure by investor type

	OO	PI	SI	LI
Purchase	0.028***	0.026***	-0.006	0.028***
	(0.008)	(0.007)	(0.006)	(0.005)
Sale	-0.060***	-0.044***	0.041**	-0.044***
	(0.016)	(0.012)	(0.012)	(0.012)
Construction	X			
New construction	X			
Holding period	X			
County FE	X			
Adj. R-sq.	0.175			
Observations	19,626,314			

This table shows regression results with annualized excess capital gains of individual repeat sales as the dependent variable. Panel A reports the average capital gains of investors relative to owner-occupiers (OO), after including different sets of one-quarter lags of "Local return dispersion" (relative to nationwide dispersion) at the county-level. Each set of lags corresponds either to the purchase quarter, the sale quarter, or both, respectively, as indicated in the row "Local return dispersion". "PI", "SI", "LI" are dummy coefficients indicating that the corresponding trade was made by a private investor, short-term institutional investor or long-term institutional investor, respectively. Private investors are defined as individuals who do not live in the underlying homes. A transaction is classified as "short-term" if the corresponding holding period is three years or less, and "long-term" otherwise. Panel B shows estimates of exposure to local return dispersion by investor type both in the purchase and sale quarter. The regression results correspond to Column (4) of Panel A. "Construction" indicates that the purchase transaction has associated documents indicating construction activity, excluding new constructions. "New construction" indicates that the purchase is identified as a transaction of a newly built home. "Holding period" stands for dummies based on holding-period sample deciles. "County FE" stands for county-level fixed-effect dummies. Standard errors are clustered over counties. *,**, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively.

posure of different investor types might be driven mainly by the upside or downside variation in local return dispersion. Therefore, we define upside potential (downside risk) as the standard deviation of above-average (below-average) annualized excess capital gains in a given county-quarter, relative to the corresponding nationwide standard deviation of above-average (below-average) annualized excess capital gains at that time. Hence, for each local market c, we compute the upside potential up_{ct} and downside risk dr_{ct} as

$$up_{ct} = \frac{1}{\tilde{N}-1} \sum_{i=1}^{\tilde{N}} \Big[max(r_{ict} - \mu_{ct}, 0)\Big]^2 \tag{4}$$

$$dr_{ct} = \frac{1}{\tilde{N}-1} \sum_{i=1}^{\tilde{N}} \Big[min(r_{ict} - \mu_{ct}, 0)\Big]^2, \tag{5}$$

with annualized excess capital gain r_{ict} and average capital gain μ_{ct} realized in quarter t, based on the number $\tilde{N}$ of capital gains, which are above or below the average return as the threshold, respectively. Again, in line with our baseline measure, we normalize both measures by dividing them by the corresponding overall upside potential and downside risk over the cross-sectional sample of all capital gains realized in period t.

Panel A of Table 4 documents the relative investor alphas under alternative channels. In Panel B, we report the factor exposure of the corresponding channels. The first two columns show that the performance differences can be explained by the upside potential as a predictor in a local market prior to the sale. Accounting for the within-county upside potential prior to a sale leads to similar performances of PIs and SIs compared to OOs. Conditional on this explanatory variable, only LIs outperform OOs, although a large proportion of their performance can be explained as well. This finding is consistent with the previous result when testing the overall return dispersion as predictor. Panel B also suggests that only institutional investors (SI as well as LI) have a positive factor loading, providing some evidence that both groups can exploit the upside potential in a local market, whereas there is no statistically significant evidence for private and owner-occupied households. We also show that within-county downside risk does not predict disparities in realized capital gains, although in general, market participants demand compensation in terms of higher capital gains, as suggested by the estimated positive and significant factor loadings. We show the estimates for investor-specific exposure at the sale date, conditional on the full set of controls. Our model specification also accounts for downside risk prior to the purchase date.

Additionally, we test alternative channels to rule them out as predictors for the SI outperformance. We study the investor-specific exposure to one-quarter lagged local average market growth to test whether return momentum can predict the observed differences in realized capital gains. For instance, several studies relate speculation in the housing market to momentum-based trading strategies (e.g., DeFusco et al., 2022; Gao et al., 2020; Bayer et al., 2020, 2021). The factor loadings are positive and statistically significant for all investor groups, which is consistent with previous studies. Comparing the coefficient size of factor exposure across investor types, we show that SIs load strongest on momentum, while trading on a long-term investment horizon appears to depend much less on recent house price movements. However, this heterogeneity in exposure does not translate into the documented capital gain disparities. Despite sizable differences in momentum exposure, the outperformance of investor groups remains at the same order of magnitude, conditional on controlling for momentum in the purchase and sale quarter. Hence, momentum trading contributes little to the outperformance of SIs, and cannot explain the disparities in realized capital gains between the different investor types.[11]

In additional robustness tests that we present in Table A3 in the Online Appendix (currently at the end of this manuscript), we also show that the capital gain disparities are unrelated to macro-financial risk factors or local economic fundamentals.[12] We include the S&P Case-Shiller US National Home Price Index to analyze the impact of the sizable housing market momentum on investor performance (e.g., Case and Shiller, 1989). Additionally, we include stock market returns and the 30-year mortgage rate to investigate whether changes in the aggregate economy impact investors differently. For instance, Peng and Zhang (2021) provide empirical evidence of heterogeneity in housing investments with respect to systematic stock market risk. Similarly, investors might differ in their borrowing costs and financing access, which could lead to a heterogeneous exposure to fluctuations in the mortgage rate. We also control for within-county local market fundamentals, such as income, unemployment, population density, and turnover as a proxy for market liquidity. We also analyze investor sensitivity to the average LTV in the local market, in order to capture lower down-payment requirements (Anenberg et al., 2019). In so doing, we mitigate potential concerns that the

[11] As suggested by Figure A4 in the Online Appendix (currently at the end of this manuscript), the persistent performance of SIs might be masked by the early 2000s housing boom, during which the outperformance relative to OOs can be misattributed to growth-driven momentum trading strategies. To reveal local return dispersion as a predictor, we show its persistence in explaining the disparities in realized capital gains conditional on the exposure to lagged market price growth in Table A2 of the Online Appendix (currently at the end of this manuscript).

[12] In each regression reported in Panel A of Table A3, we interact the returns of the variable given in the first column with each buyer identity. As each repeat sale is exposed to each factor at the purchase as well as the sale date, respectively, we include realizations of the given variable at both points in time. For all of the models analyzed, the estimated investor-specific factor loadings remain positive at the same order of magnitude and highly significant. For the results in Panel B, we instead transform the variables to annualized changes over the respective holding periods, similar to the holding-period factor models proposed by Peng (2016) and Peng and Zhang (2021). Again, the outperformance of investors remains at the same order of magnitude and is highly significant for all specifications tested.

Table 4
Testing upside potential and alternative channels

Panel A: Relative investor alphas under alternative channels

	Upside potential	Upside potential	Downside risk	Momentum
PI	0.004	0.004	0.010**	0.015***
	(0.005)	(0.005)	(0.004)	(0.001)
SI	-0.000	-0.002	0.082***	0.048***
	(0.014)	(0.015)	(0.014)	(0.003)
LI	0.011*	0.016***	0.032***	0.037***
	(0.005)	(0.005)	(0.006)	(0.001)
Channel	sale	purchase/sale	purchase/sale	purchase/sale
Construction	X	X	X	X
New construction	X	X	X	X
Holding period	X	X	X	X
County FE	X	X	X	X
Adj. R-sq.	0.164	0.171	0.174	0.206
Observations	20,900,378	19,442,042	19,237,803	20,656,495

Panel B: Corresponding exposure to channel (at sale) by investor type

	Upside potential	Upside potential	Downside risk	Momentum
OO	-0.005	-0.005	0.033**	0.651***
	(0.007)	(0.009)	(0.011)	(0.044)
PI	0.004	0.004	0.048***	0.623***
	(0.007)	(0.008)	(0.012)	(0.040)
SI	0.045***	0.047***	0.013	0.807***
	(0.012)	(0.012)	(0.009)	(0.048)
LI	0.015**	0.014*	0.027**	0.267***
	(0.006)	(0.006)	(0.010)	(0.018)
Exposure at purchase	-	X	X	X
Construction	X	X	X	X
New construction	X	X	X	X
Holding period	X	X	X	X
County FE	X	X	X	X
Adj. R-sq.	0.164	0.171	0.174	0.206
Observations	20,900,378	19,442,042	19,237,803	20,656,495

This table shows regression results with annualized excess capital gains as the dependent variable. Panel A reports the average capital gains of investors relative to owner-occupiers (OO), after including one-quarter lags of "Upside Potential", "Downside Risk", and "Momentum" at the county-level. Each set of lags corresponds either to the purchase quarter, sale quarter, or both. We define upside potential (downside risk) as the standard deviation of above-average (below-average) annualized excess capital gains in a given county-quarter, relative to the corresponding nationwide above-average (below-average) annualized excess returns in that quarter. Momentum is defined as the county-level index return. "PI", "SI", "LI" are dummy coefficients indicating that the corresponding trade was made by a private investor, short-term institutional investor or long-term institutional investor, respectively. A transaction is classified as "short-term" if the corresponding holding period is three years or less, and "long-term" otherwise. Panel B shows the corresponding factor exposure by investor type at the sale quarter. "Construction" indicates that the purchase transaction has associated documents indicating construction activity, excluding new constructions. "New construction" indicates that the purchase is identified as a transaction of a newly built home. "Holding period" stands for dummies based on holding-period sample deciles. "County FE" stands for county-level fixed-effect dummies. Standard errors are clustered over counties. *,**, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively.

performance differences can be generally explained by higher local exposure to debt-driven housing markets. Neither of the tested variables can help explain the disparities in investor performance. Our results suggest that the investor groups are not systematically rewarded with higher compensating returns for targeting markets in which they are exposed differently to local macroeconomic fundamentals. Also, they do not systematically gain more from changes in local macroeconomic conditions when timing their investments in particular states of the housing market cycle.

3.5 Which housing markets do investors enter?

In the previous section, we demonstrate that the within-county exposure to local return dispersion prior to a sale can explain differences in capital gains across investor groups. We also identify local upside potential in the market as a predictor of the outperformance of SIs. Their exposure to this upside variation in return dispersion is likely to be related to the selection process underlying the purchase decision. This decision should be linked to the investor's underlying strategy, and potentially has a high impact. In this section, we thus investigate potential determinants of market selection across investor groups, in particular local return dispersion and momentum trading.

We do so by forming indicator variables that equal one if a purchase was made by a particular group of investors and zero otherwise. We then run probit regressions to investigate the relationship between investor group activity and lagged local return dispersion in addition to control variables such as income or unemployment at the time of purchase. Accordingly, we model the conditional probability of buyer i belonging to group k to purchase a home in county c at time $T = t_1$ as

$$P(Buyer\ i\ belongs\ to\ group\ k|X_{ct_1-1}, T = t_1) = \Phi(X_{ct_1-1}\beta + \nu_{t_1}), \tag{6}$$

in which $\Phi()$ is the normal cumulative probability distribution, X_{ct_1-1} a set of lagged explanatory variables, β a vector of parameters, ν_{t_1} captures purchase-quarter fixed-effect dummies, and the set of groups is $k \in \{\text{OO, PI, SI, LI}\}$. For each type of investors, we run a separate regression to investigate which investor groups buy in higher-risk locations, and whether past information on return dispersion matters as well. In matrix X_{ct_1-1}, we also include local momentum to determine at which state of the local housing market cycle investors are particularly active, and to account for the effect of potential trading on autocorrelation. For all regressors, we use quarter-lagged variables to further mitigate simultaneity issues. All variables are measured at the county level. The included quarter-year dummies additionally filter out common macroeconomic determinants and allow for a cross-sectional comparison.

Table 5 reports the results of four probit regressions on the probability of a given investor group making a particular purchase. The first column investigates the activity of owner-occupied households (OO) relative to all other investor types. OOs tend to invest in counties with lower return dispersion, as suggested by its negative and significant coefficient estimate. Thus, a higher experienced return uncertainty seems to reduce their purchase activity. The coefficient for the index return is negative as well, suggesting that OOs tend to buy in local markets with falling prices, relative to the other investor groups. Furthermore, purchase activity tends to be high when unemployment is low. This could be linked to owner-occupied households preferring locations close to their workplace, or that homes become more affordable to a wider range of individuals when employment is high. Lastly, a predictor of OO activity is the average primary LTV ratio, which could indicate that owner-occupied households benefit from easy access to credit, or are willing to take out larger loans to buy a home in the preferred location.

The results for the other investor types have to be interpreted largely relative to OOs, who make up about 75% of the total observations. All three investor groups, PIs, SIs, and LIs, are more likely to invest in counties with comparably high return dispersion. This could reflect a general rent-seeking pattern, in which investors aim for higher expected capital gains by engaging in riskier trades to seek the upward potential when selling. By contrast, the importance of local momentum varies across investor types. As indicated by the large positive and highly significant coefficient, PIs are mostly active in booming markets. SIs tend to invest in falling markets, which might explain their high capital gains realized in the post-bust recovery phase of the housing market cycle, allowing them to exploit the upward potential.

The LI activity is not significantly related to momentum, suggesting that current price movements lose relevance for long-term strategies of professionals. This result is intuitive, as short-term deviations from the general house price trend should not matter if homes are held over the long-term. Our finding suggests that LIs indeed target long-term holding periods when buying the property, underpinning our cut-off choice of three years. Other than that, institutional investors choose similar locations, characterized by higher population density, lower income, and lower unemployment. This could suggest that investors tend to buy homes in gentrifying neighborhoods. In summary, all investors, especially SIs and PIs, tend to invest in markets with relatively high return dispersion. When it comes to local momentum, however, strategies appear to be different. PIs seem to be active mostly in booming markets, while SIs tend to buy in falling ones. In contrast to all other groups, LIs do not vary their activity with local price trends.

Our next step is to study whether the SI outperformance can be linked to a specific

Table 5
Local variables and the purchase decision of investor groups

	OO	PI	SI	LI
County return dispersion	-0.319***	0.280**	0.251***	0.147***
	(0.084)	(0.091)	(0.026)	(0.030)
County momentum	-0.891**	0.976**	-0.175*	-0.021
	(0.290)	(0.311)	(0.085)	(0.096)
Log county pop. density	0.002	-0.019	0.063***	0.037***
	(0.026)	(0.027)	(0.007)	(0.010)
Log county income	0.223	-0.163	-0.211***	-0.300***
	(0.154)	(0.166)	(0.042)	(0.044)
County unemployment	-0.044**	0.050***	-0.010*	-0.006
	(0.014)	(0.015)	(0.005)	(0.005)
Log county turnover	0.025	-0.017	-0.025**	-0.028*
	(0.024)	(0.025)	(0.009)	(0.011)
County primary LTV	1.233***	-1.078***	-0.579***	-0.761***
	(0.146)	(0.133)	(0.058)	(0.077)
Purchase quarter FE	X	X	X	X
Pseudo R-sq.	0.041	0.035	0.062	0.030
Observations	19,206,741	19,206,741	19,206,741	18,873,506

This table shows the results of four probit regressions that investigate whether a particular investment was made by one of our four investor groups, owner-occupiers (OOs), private investors (PIs), short-term institutional (SI), or long-term institutional investors (LI), respectively. Owner-occupants are private individuals who live in the home underlying the transaction. Private investors are defined as individuals who do not live in the respective home. A transaction is classified as "short-term" if the corresponding holding period is three years or less, and "long-term" otherwise. All regressors are lagged variables, i.e., quarterly measured variables are one-quarter lags and yearly measured variables (income, unemployment and population density) are yearly lags. "County return dispersion" indicates the county-quarter level return dispersion in the purchase quarter of the underlying transaction. Return dispersion is measured as local relative to nationwide dispersion of annualized capital gains within the same quarter. "County momentum" indicates the county-quarter level index return in the purchase quarter of the underlying transaction. "Log county pop. density" is the natural log of the county's population density, derived as number of inhabitants in the purchase year per square-mile of land. "Log county income" is the natural log of the median county income in the purchase year, "County unemployment" is the county's unemployment rate in percentage points in the purchase year, "Log county turnover" the natural log of the sum of total transactions in the purchase quarter, and "County primary LTV" is the average mortgage amount divided by the average sales price in a given county in a given quarter, excluding LTVs larger than one. All regressions include purchase-year-quarter fixed-effect dummies. Standard errors are clustered over counties. *,**, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively.

trading strategy. While Table 5 suggests that SIs are particularly more active in markets with high return dispersion, it remains unclear whether the uncovered pattern, combined with the outperformance in such markets, is indeed evidence of a generalizable strategy. As suggested by Table 3, SIs benefit from capital gains uncertainty at the time of sale, conditional on being already invested in a market. However, as local return dispersion prior to the location choice does not predict the disparities in realized capital gains within a market, we also focus on return dispersion growth. Since SIs target a short-term investment horizon, observed dispersion growth before market entry might serve as a additional predictor of potential capital gains realized over a short holding period. Therefore, we hypothesize SIs to anticipate such markets on purchase, and the performance differences should be explained by this return dispersion pattern prior to their investment decision.

In Table 6, we analyze such market selection behavior. Panel A documents the average activity of SIs in a given county relative to nationwide short-term activity for different quantiles of return dispersion and its growth rate, respectively. To be able to interpret the results as actual selection behavior, we calculate the corresponding quantiles dynamically, based on the year prior to the purchase year. That is, for each (lagged) year, we divide all counties into ten return-dispersion quantiles. We proceed analogously for the growth rate of return dispersion, however, by dividing it into only three categories: low growth (1st decile), average growth (2nd to 9th quantile), and high growth (10th quantile). In line with our intuition, we observe the highest SI activity in highly dispersed markets with high dispersion growth. We confirm the results for the full sample, the boom period from 1995 to 2006, and the subsequent bust and recovery phase from 2007 to 2017. The findings indicate that SIs indeed tend to dynamically select markets, based on a return dispersion pattern. Although the pattern is stronger during the 1995-2006 period, it still holds remarkably well for the years 2007-2017, suggesting a persistent strategy over time.

Panel B of Table 6 shows the corresponding performance results based on the return-dispersion pattern prior to the purchase. The regressions displayed are similar to our base case ones, but include our dispersion measures using dummies rather than continuous variables to reflect that we are measuring dispersion using quantiles. Hence, the relative performances must be interpreted with respect to a base category, which is the low-dispersion, low-growth intersection. The insignificant coefficients for the SI indicate that, as before, in locations with little return dispersion, SI do not outperform OOs, which holds for the full sample as well as the two subperiods analyzed. Interestingly, in such markets with low dispersion, LIs appear to perform quite well, suggesting the long-term strategy pays off particularly in calmer markets. More important, capturing local return dispersion and its growth prior to the purchase can explain the SI outperformance. Having shown that SIs seem to actively select highly dispersed locations, in the next section, we test a potential strategy based on the purchase of distressed properties, to exploit return dispersion, and specifically the upside potential in such markets.

3.6 Do short-term institutional investors bet on upside potential?

Section 3.4 documents that the upside variation of the local return dispersion prior to selling the property predicts the outperformance of SIs. However, this raises the question of whether SIs persistently bet on the upside potential in local markets. If this is the case, they should specifically target markets in which they can later obtain higher realized capital gains when selling the property. In this section, we therefore test whether higher SI activity can be related to local markets with higher right-skewed variation in capital gains realized upon sale. To mitigate potential endogeneity concerns, we adapt an instrumental variable strategy.

A possible trading strategy through which SIs can exploit the upside potential of the local return dispersion might be to strategically buy distressed properties. For instance, Lambie-Hanson et al. (2022) provide empirical evidence that institutional investors to a large extent purchased foreclosure-related properties during the housing market bust.[13] Following our intuition, we hypothesize that this strategy led SIs to be more active in areas with financially distressed properties. To test this channel empirically, we follow Gupta (2019). He exploits the unexpected increase in the spread between LIBOR and Treasury rates in 2008 and 2009, caused by overnight risk, to identify local foreclosure spillovers from LIBOR indexed ARM borrowers, being more vulnerable than holders of Treasury-rate-related ARMs. We thus expect areas with an exceptionally high 2007 share of LIBOR relative to Treasury linked ARMs to be associated with an increased activity of SIs, whose trading strategy seems to rely on buying such distressed properties.

As instrument for SI activity, we use the difference between LIBOR and Constant Maturity Treasury (CMT)-linked ARMs relative to the total transactions observed in a given area in 2007, $\Delta LC_{2007} = \frac{\Delta LIB - \Delta CMT}{Transactions_{Total}}$. For the exclusion restriction to be fulfilled, we assume that the instrument does not directly affect the upside potential in a local market other than through SI activity. First, the spread between LIBOR and Treasury rates had been rather steady until 2007 (Gupta, 2019). Therefore, the financing choice between ARMs linked either to the LIBOR rate or CMTs was thus likely to be random prior to the bubble

[13] To further motivate our intuition, we also refer to Figure A6 in the Online Appendix (currently at the end of this manuscript) for additional support. The graph shows that during the housing bust, SIs increasingly purchased foreclosed or REO-related properties (up to 50%), which are typically subject to large price discounts (e.g., Campbell et al. (2011)). We observe a similar pattern prior to the early 2000s boom period. This might explain the higher capital gains earned by SIs after investing in distressed market segments.

Table 6
SI activity and performance by local return dispersion and dispersion growth

Panel A: SI purchase activity, dynamically sorted by prior year's local return dispersion and growth rate

	Local dispersion growth rate								
	Full sample			1995-2006			2007-2017		
Level local return dispersion	Low	Average	High	Low	Average	High	Low	Average	High
Low (1)	0.041	0.043	0.048	0.017	0.028	0.044	0.048	0.049	0.051
2	0.040	0.050	0.055	0.028	0.035	0.048	0.047	0.056	0.059
3	0.050	0.058	0.055	0.043	0.047	0.047	0.054	0.062	0.059
4	0.032	0.059	0.057	0.025	0.050	0.054	0.036	0.063	0.059
5	0.044	0.065	0.069	0.032	0.059	0.063	0.048	0.068	0.071
6	0.041	0.068	0.094	0.031	0.058	0.069	0.048	0.073	0.101
7	0.042	0.074	0.084	0.031	0.063	0.082	0.050	0.080	0.085
8	0.054	0.076	0.117	0.039	0.060	0.150	0.064	0.084	0.110
9	0.063	0.083	0.116	0.047	0.063	0.117	0.068	0.093	0.116
High (10)	0.080	0.109	0.167	0.079	0.098	0.198	0.081	0.114	0.155

Panel B: Corresponding relative investor performance in low/low quantiles

	Full sample	1995-2006	2007-2017
PI	0.006	0.011*	0.009*
	(0.004)	(0.005)	(0.004)
SI	-0.003	-0.001	0.022
	(0.009)	(0.009)	(0.012)
LI	0.038***	0.047***	0.027***
	(0.004)	(0.013)	(0.004)
Controls	yes	yes	yes
Dispersion growth quantiles (dynamic)	yes	yes	yes
Dispersion level quantiles (dynamic)	yes	yes	yes
Adj. R-sq.	0.171	0.189	0.133
Observations	19,447,113	6,304,442	13,142,671

Panel A of this table reports the average county short-term institutional investor (SI) activity, measured by the share of SIs in a given county relative to nation-wide short-term activity. These averages are divided by quantiles of local dispersion growth rate (given in the columns; low: lowest decile, average: second to ninth decile, high: highest decile) and the local dispersion level (given in the rows). The quantiles are derived dynamically, i.e., calculated from the quantile of a county in the year prior to the respective purchase transaction. The results are given for the full sample period, the boom period from 1995 to 2006 and the subsequent period from 2007 to 2017. Panel B shows the performance of investors relative to owner-occupiers (OO) when including the quantiles from Panel A, interacted with each investor type. "PI", "SI", "LI" are dummy coefficients indicating that the corresponding trade is made by a private investor, short-term institutional investor or long-term institutional investor, respectively. All regressions include the same set of control variables: Dummies for construction, new construction, and holding-period sample deciles, as well as county-level fixed effects. Standard errors are clustered over counties. *,**, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively.

burst, which validates this variable as an instrument for our empirical analysis. Second, our regression approach is based on a time lag between the instrument, which is observed prior to SIs investment timing, and potential right-skewed variation in realized capital gains upon sale. We expect larger SI activity in areas with high LIBOR ARM exposure, relative to the CMT ARM exposure (with the latter controlling simultaneously for the general ARM activity). Higher purchase activity is then followed naturally by the subsequent increase in SI sales. In particular, as the ARM shock was observed in 2008 and 2009, we expect the increasing distress and thus purchase activity to start in 2010, once interest payments had been adjusted, and to continue in the following years. As properties are typically held for one or two years (e.g., due to tax incentives), we expect sales activity to increase in 2011 and the years to follow.

We proceed with a standard 2SLS approach. First, we regress the share of SI sales activity in a county relative to nationwide short-term trading activity on our instrument. Our first stage regression equals

$$SI_{c,t} = \alpha + \beta \Delta LC_{2007} + \gamma X_{c,t-} + \nu_t + \varepsilon_{c,t}, \tag{7}$$

in which we also control for local fundamentals observed in 2007 as well as annual time fixed-effects. We use pre-treatment variables to avoid control variables being affected by the treatment as well, e.g., through feedback effects. In the second stage, we then regress local return dispersion on our instrumented SI activity, including the same set of control variables:

$$D_{c,t} = \alpha + \beta \widehat{SI}_{c,t} + \gamma X_{c,t-} + \nu_t + \varepsilon_{c,t}. \tag{8}$$

Table 7 shows first and second stage regression results for the SI sales activity for the years following the 2009 shock to LIBOR ARMs. In this analysis, the sample ranges from 2009 to 2014, to account for the typical short-term holding period after the purchase. As we expect purchase activity to increase directly after 2009, by construction, sales activity increases with a certain delay in 2013 and 2014, respectively. We also restrict the sample to counties with LIBOR exposure in 2007 of at least 5%, i.e., at least 5% of transactions in 2007 in a given area should be financed with LIBOR-indexed ARMs to ensure a sufficient degree of treatment.[14]

In our specification, we use the time-invariant instrument as well as a more flexible structure, allowing for a heterogeneous effect each year. The overall ΔLIB-ΔCMT instrument is positive and statistically significant at the 5% level. Interacting the instrument with each

[14]The results hold when we only include counties with a higher share of LIBOR-based ARMs or additionally require a positive share of CMT ARMs.

year, the SI sales activity increases overproportionally with a delay of the short-term holding period in LIBOR-ARM-exposed areas in the years after the 2009 shock. The results hold, conditional on year fixed effects and multiple controls, including the 2007 SI activity, macroeconomic fundamentals, as well as return dispersion and momentum in that year. County-level fixed effects are not included, as our instrument is time-invariant by construction and derived at the same aggregation level.

For both the time-invariant and the by-year models, we observe a positive and statistically significant cross-sectional relationship between the instrumented SI activity and return dispersion in local markets. Testing the causal effect on the upside potential, we also find a large, positive and statistically significant coefficient, suggesting that a larger share of SIs in a local market tends to increase the upside potential of realized capital gains. The findings are in line with our intuition that SIs can exploit the upside potential when purchasing distressed properties in a local market. As a counterfactual test, we also show that the share of SIs has no statistically significant effect on the downside variation of local return dispersion.

4 Robustness tests

In Section 3, we document that performance differences between investors can be explained by local return dispersion, i.e., we test local return dispersion as a predictor for capital gains. The aim of this section is to provide further evidence that this result is robust to alternative model specifications and subsamples. Table 8 summarizes our robustness checks.

Panel A of Table 8 reports the relative alphas of investors from regression models with alternative versions of local return risk as predictor. First, following Han (2013), we address issues with simultaneity and look-ahead bias by instrumenting local return dispersion with one and two of its lags, respectively. Using both of these instruments, we find that the outperformance of investor groups relative to OOs can be fully explained. We also show that the results do not change when we use a one-year lag of local return dispersion to capture potential time restrictions and delays between investor trading decisions and the actual sale.

Our full sample contains repeat sales generated between 1995 and 2017. Hence, at the beginning of our sample, we observe capital gains realized over very short holding periods. To ease this restriction, we investigate two different subsamples that allow for a "burn-in" period of five years. We show the results when investigating transactions with sale dates between 2000 and 2015 only, but allowing purchases to range back to 1995. For this subperiod, the estimated alphas become insignificant when we include local return dispersion as the explanatory variable. Similarly, we additionally test a subsample with transactions realized between 2000 and 2010, for two reasons. First, by investigating a period of 11 years while

Table 7
Betting on upside potential: Instrumenting SI sales activity

	Instrumented variables		Local return dispersion		Upside potential		Downside risk	
	1st stage	1st stage (by year)	2nd stage	2nd stage (by year)	2nd stage	2nd stage (by year)	2nd stage	2nd stage (by year)
ΔLIB-ΔCMT (2007)	0.200**							
	(0.075)							
SI seller share (IV)			4.308**	3.039***	4.960**	4.125***	0.508	0.690
			(1.462)	(0.828)	(1.718)	(1.026)	(1.012)	(0.647)
ΔLIB-ΔCMT (2007)×2009		-0.308***						
		(0.058)						
ΔLIB-ΔCMT (2007)×2010		-0.255***						
		(0.060)						
ΔLIB-ΔCMT (2007)×2011		-0.145**						
		(0.047)						
ΔLIB-ΔCMT (2007)×2012		0.022						
		(0.050)						
ΔLIB-ΔCMT (2007)×2013		0.144**						
		(0.054)						
ΔLIB-ΔCMT (2007)×2014		0.374***						
		(0.110)						
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-sq.	0.328	0.385	0.593	0.598	0.416	0.453	0.416	0.419
Observations	485,642	485,642	483,075	483,075	482,485	482,485	481,856	481,856

This table shows first- and second-stage regression results of our 2SLS instrumental variable approach. Columns (1) and (2) show the first-stage results of the county-level share of short-term institutional investor (SI) activity, relative to nationwide share of short-term sales, regressed on the time-invariant instrument (ΔLIB-ΔCMT) and its heterogeneous effect each year, based on interactions with year dummies. Instrumental variable "ΔLIB-ΔCMT" is defined as the 2007 share of LIBOR indexed ARMs relative to overall transactions minus the same ratio for Constant Maturity Treasury (CMT)-based ARMs that year in a given county. Columns (3) to (8) show second-stage cross-sectional regression results of local return dispersion, upside potential, and downside risk (all relative to their nationwide levels), respectively, regressed on the instrumented share of SI activity. The sample ranges from 2009 to 2014. Control variables include 2007 county-levels of the share of SI activity, local return dispersion, momentum, log income, unemployment rate and population density, preceding the unexpected LIBOR-CMT spread in 2008/2009. Standard errors are clustered over counties. *,**, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively.

allowing for a 5-year burn-in period, we compare observations with more similar holding periods. Second, this period covers the most recent housing boom and bust cycle, allowing us to analyze the period with the highest reward, but also the highest risk. The results indicate that during this period, none of the investors outperforms OOs after adjusting for local return dispersion. In other words, the insignificant coefficients confirm local return dispersion as a predictor for annualized excess capital gains.

In Panel B, we test alternative investor definitions. In the baseline regressions, we specify a holding period of three years as the cut-off for short-term trading, following DeFusco et al. (2022). As this might be considered a rather subjective choice, we first restrict all groups to have a holding period of three or less years. We do so to underscore that our holding-period controls perform well in removing short- and long-term-related activities. Our results show that when comparing short-term investors only, investor outperformance relative to OOs can also be predicted by local return dispersion. Second, we test whether our results still hold when using two or four years as a breakpoint instead. For instance, Bayer et al. (2020) show that 90% of large-scale flippers resell homes within four years after the purchase. Our findings suggest that when using two years as the cut-off, the outperformance of all three investor types remains insignificant, whereas using four years leads to a significant outperformance of LIs. Thus, our key result that lagged local return dispersion predicts short-term investor outperformance is robust to alternative cut-off values. Given that the outperformance of LIs seems to emerge with a breakpoint between two and four years, our choice of three years appears to be a balanced one. Finally, we show that local return dispersion explains the performance differences when we propose an alternative definition to distinguish between SIs and LIs. In this robustness test, companies are first identified by coinciding buyer names. We then assign SI (LI) identity to all transactions of a company if at least (less than) 75% of its trades are made within three years. In sum, our robustness checks provide supporting evidence of local return dispersion as an important predictor of systematic differences in capital gains across investor groups.

5 Conclusion

In this paper, we study the substantial and persistent performance differences in realized capital gains among different types of investors in the housing markets. These groups of investors can be expected to differ in their underlying strategies and primary goals. While OOs should be primarily interested in drawing utility from housing consumption, investors are likely to focus on financial interests in their homes. We thus expect investors to require compensation for engaging in riskier trades. Although PIs sometimes invest in second homes

Table 8
Robustness checks: Relative performance conditional on local return dispersion

Panel A: Alternative variable definition and subsamples

	IV 1 lag	IV 2 lags	Yearly disp.	2000-2015	2000-2010
PI	-0.002	-0.002	-0.005	-0.005	-0.012
	(0.007)	(0.007)	(0.005)	(0.006)	(0.007)
SI	-0.022	-0.029	-0.005	-0.026	-0.027*
	(0.019)	(0.019)	(0.014)	(0.014)	(0.013)
LI	0.011	0.013	0.005	0.012	0.001
	(0.008)	(0.008)	(0.006)	(0.007)	(0.011)
Adj. R-sq.	0.175	0.179	0.179	0.186	0.159
Observations	19,626,314	19,058,465	19,310,981	15,933,605	9,705,531

Panel B: Alternative investor definition

	HP < 3 years	ST 2 yr	ST 4 yr	Alt. SI
PI	0.011	0.001	0.001	0.001
	(0.008)	(0.005)	(0.005)	(0.005)
SI	0.014	-0.016	-0.009	0.013
	(0.015)	(0.017)	(0.014)	(0.016)
LI		0.010	0.019**	-0.002
		(0.007)	(0.006)	(0.007)
Adj. R-sq.	0.115	0.175	0.175	0.175
Observations	5,994,532	19,626,314	19,626,314	19,626,314

This table shows the results of OLS regressions with annualized excess capital gains as the dependent variable. The displayed coefficients are estimates for group-specific dummies. The reference group consists of owner-occupiers (OO), i.e., private individuals who live in the home purchased during the holding period. "PI", "SI", "LI" are dummy coefficients indicating that the corresponding trade is made by a private investor, short-term institutional investor or long-term institutional investor, respectively. Private investors are defined as individuals who do not live in the underlying purchased home. A transaction is classified as "short-term" (ST) if the corresponding holding period is three years or less, and "long-term" otherwise. If not indicated otherwise, the regressions for which results are displayed in this table include group-specific exposure to the one-quarter lag of county return dispersion in the purchase and sale quarter, respectively. All regressions include controls for "Construction", "New construction", and "Holding period" deciles, as well as county-level fixed effects. Panel A shows regression results when using its one-period and two-period quarter lags as instruments for contemporaneous local return dispersion, similar to Han (2013), as well as when using the one-year lag of local return dispersion instead of the quarterly measure, or using subsamples based on transactions realized between 2000-2015 as well as 2000-2010, respectively. Panel B shows the results when using only transactions with a holding period of less than three years, altering the definition of short-term trading (ST) to either two years or four years, respectively, and defining SI (LI) as investors with at least (less than) 75% of trades made within three years. *,**, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively.

to save for retirement, their activity increased during the last housing boom, suggesting tendencies to speculate on momentum. Short-term institutional investors (SIs), in turn, are likely to engage in speculative activity, aiming for high short-term returns on their investments. Long-term institutional investors (LIs), in contrast, should rather trade for steady dividends from rental income. Varying in their primary motivation and their underlying trading strategies, investors and OOs are likely to be heterogeneously exposed to risk and momentum, affecting their respective performance.

Using nationwide data on repeat sales of US residential real estate, we document sizable and persistent differences in annualized excess capital gains among buyers in the housing market. We find that the observed performance differences across investors in the same local market can be explained by their heterogeneous exposure to past local return dispersion. By contrast, neither selection of particular locations based on macroeconomic fundamentals, nor timing with respect to the aggregate or local housing market states, can explain the performance differences among market participants.

We further document that the outperformance of SIs can be predicted by the upside variation in local return dispersion. Building on these new insights raises the question of whether SIs can exploit a persistent strategy to trade on the upside potential in local markets. Applying an instrumental variable approach, we confirm that SI activity tends to be positively related to the upside variation in local return dispersion in markets in which they purchased distressed properties. Hence, our findings shed new light on housing market speculation, offering new evidence of betting on the upside potential as an investment strategy to outperform other market participants.

References

ANENBERG, E., A. HIZMO, E. KUNG, AND R. MOLLOW (2019): "Measuring Mortgage Credit Availability: A Frontier Estimation Approach," *Journal of Applied Econometrics*, 34, 865–882.

ANENBERG, E. AND E. KUNG (2014): "Estimates of the Size and Source of Price Declines Due to Nearby Foreclosures," *American Economic Review*, 104, 2527–2551.

BAYER, P., C. GEISSLER, K. MANGUM, AND J. W. ROBERTS (2020): "Speculators and Middlemen: The Strategy and Performance of Investors in the Housing Market," *Review of Financial Studies*, 33, 5212–5247.

BAYER, P., K. MANGUM, AND J. W. ROBERTS (2021): "Speculative Fever: Investor Contagion in the Housing Bubble," *American Economic Review*, 111, 609–651.

BLACK, F., M. C. JENSEN, AND M. S. SCHOLES (1972): "The Capital Asset Pricing Model: Some Empirical Tests," in *Studies in the Theory of Capital Markets*, ed. by M. C. Jensen, Praeger Publishers Inc.

BOLLERSLEV, T., A. J. PATTON, AND W. WANG (2016): "Daily House Price Indices: Construction, Modeling, and Longer-Run Predictions," *Journal of Applied Econometrics*, 31, 1005–1025.

CAMPBELL, J. Y. AND J. F. COCCO (2007): "How Do House Prices Affect Consumption? Evidence from Micro Data," *Journal of Monetary Economics*, 54, 591–621.

CAMPBELL, J. Y., S. GIGLIO, AND P. PATHAK (2011): "Forced Sales and House Prices," *American Economic Review*, 101, 2108–2131.

CASE, K. E. AND R. J. SHILLER (1989): "The Efficiency of the Market for Single-Family Homes," *American Economic Review*, 79, 125–137.

CHINCO, A. AND C. MAYER (2016): "Misinformed Speculators and Mispricing in the Housing Market," *Review of Financial Studies*, 29, 486–522.

CORRADIN, S. AND A. POPOV (2015): "House Prices, Home Equity Borrowing, and Entrepreneurship," *Review of Financial Studies*, 28, 2399–2428.

CVIJANOVIC, D. AND C. SPAENJERS (2021): "'We'll Always Have Paris': Out-of-Country Buyers in the Housing Market," *Management Science*, 67, 4120–4138.

DeFusco, A., C. Nathanson, and E. Zwick (2022): "Speculative Dynamics of Prices and Volume," NBER Working Paper No. w23449, Available at SSRN: https://ssrn.com/abstract=2976189.

Fischer, M., R. Füss, and S. Stehle (2021): "Local House Price Comovements," *Real Estate Economics*, 49, 169–198.

Fu, Y. and W. Qian (2014): "Speculators and Price Overreaction in the Housing Market," *Real Estate Economics*, 42, 977–1007.

Fu, Y., W. Qian, and B. Yeung (2016): "Speculative Investors and Transactions Tax: Evidence from the Housing Market," *Management Science*, 62, 3254–3270.

Fuster, A. and J. Vickery (2015): "Securitization and the Fixed-Rate Mortgage," *Review of Financial Studies*, 28, 176–211.

Gao, Z., M. Sockin, and W. Xiong (2020): "Economic Consequences of Housing Speculation," *Review of Fiancial Studies*, 33, 5248–5287.

Gerardi, K., E. Rosenblatt, P. S. Willen, and V. Yao (2015): "Foreclosure Externalities: New Evidence," *Journal of Urban Economics*, 87, 42–56.

Gupta, A. (2019): "Foreclosure Contagion and the Neighborhood Spillover Effects of Mortgage Defaults," *Journal of Finance*, 74, 2249–2301.

Han, L. (2010): "The Effects of Price Risk on Housing Demand: Empirical Evidence from US Markets," *Review of Financial Studies*, 23, 3889–3928.

——— (2013): "Understanding the Puzzling Risk-Return Relationship for Housing," *Review of Financial Studies*, 26, 877–928.

Hoechle, D., M. Schmid, and H. Zimmermann (2020): "Does Unobservable Heterogeneity Matter for Portfolio-Based Asset Pricing Tests?" Available at SSRN: https://ssrn.com/abstract=3569485.

Lambie-Hanson, L., W. Li, and M. Slonkosky (2022): "Real Estate Investors and the US Housing Recovery," *Real Estate Economics*, forthcoming.

Loutskina, E. and P. E. Strahan (2015): "Financial Integration, Housing, and Economic Volatility," *Journal of Financial Economics*, 115, 25–41.

Mills, J., R. Molloy, and R. Zarutskie (2019): "Large-Scale Buy-to-Rent Investors in the Single-Family Housing Market: The Emergence of a New Asset Class," *Real Estate Economics*, 47, 399–430.

Nathanson, C. G. and E. Zwick (2018): "Arrested Development: Theory and Evidence of Supply-Side Speculation in the Housing Market," *Journal of Finance*, 73, 2587–2633.

Peng, L. (2016): "The Risk and Return of Commercial Real Estate: A Property Level Analysis," *Real Estate Economics*, 44, 555–583.

Peng, L. and T. G. Thibodeau (2017): "Idiosyncratic Risk of House Prices: Evidence from 26 Million Home Sales," *Real Estate Economics*, 45, 340–375.

Peng, L. and L. Zhang (2021): "House Prices and Systematic Risk: Evidence from Microdata," *Real Estate Economics*, 49, 1069–1092.

Saiz, A. (2010): "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, 125, 1253–1296.

Online Appendix

A Additional tables and figures

Table A1
Data-cleaning process

	Dismissed observations	Remaining observations
Starting observations	-	99,757,949
Duplicates	13,790,976	85,966,973
Extreme prices	325,917	85,641,056
Resulting repeat sales	-	33,280,346
Holing period < 180 days	2,544,614	30,735,732
Extreme annualized capital gains	986,917	29,748,815
Unknown identity	3,531,577	26,217,238
Full sample extreme capital gains	524,344	25,692,894
REO/foreclosure related	4,514,025	21,178,869

This table shows the number of observations lost in each cleaning step of the housing data. The steps are followed in descending order. The upper panel shows numbers for individual transactions, the lower panel for the resulting repeat sales. The starting observations are transactions of residential homes purchased from 1995 to 2017 that are declared "arms length" by the data provider, have a positive sales price and a full date including day, month, and year of sale. Duplicates are removed according to DeFusco et al. (2022). Extreme prices (annualized capital gains) are defined as outside the 5,000 USD to 100,000,000 USD (-50%, +100%) interval, as in Bollerslev et al. (2016). Observations with "Unknown identity" are repeat sales to which no investor type could be assigned due to missing information on the traded home's, or the owner's mailing address. "Full sample extreme capital gains" are defined as exceeding the first and last percentile of the return distribution in the full sample at the given cleaning step. "REO/foreclosure related" are repeat sales for which either purchase or sale transaction was associated with a foreclosure or foreclosure-related transfer of the underlying home, respectively.

Table A2
Investor performance, local return dispersion and local momentum

	(1)	(2)	(3)	(4)
PI	0.013***	0.011**	0.005	0.006
	(0.001)	(0.004)	(0.005)	(0.005)
SI	0.049***	0.045***	-0.015	-0.010
	(0.003)	(0.008)	(0.015)	(0.015)
LI	0.031***	0.023***	0.016**	0.019***
	(0.001)	(0.005)	(0.006)	(0.006)
Local dispersion & momentum	-	purchase	sale	purchase & sale
Construction	X	X	X	X
New construction	X	X	X	X
Holding period	X	X	X	X
County FE	X	X	X	X
Adj. R-sq.	0.161	0.172	0.209	0.222
Observations	21,178,869	19,611,315	20,680,566	19,369,779

This table shows the results for regressions with annualized excess capital gains of individual repeat sales as the dependent variable. The coefficients displayed indicate average capital gains of investors relative to owner-occupiers, after including different sets of one-quarter lags of "Local momentum", i.e., the quarterly return of the county-level index, and "Local dispersion", i.e., the dispersion of county-level annualized capital gains relative to nationwide dispersion of returns within the same quarter. Each set of lags is included either relative to the purchase quarter, the sale quarter, or both, respectively, as indicated in the row "Local dispersion & momentum". "PI", "SI", "LI" are dummy coefficients indicating that the corresponding trade was made by a private investor, short-term institutional investor or long-term institutional investor, respectively. Private investors are defined as individuals who do not live in the underlying homes. A transaction is classified as "short-term" if the corresponding holding period is three years or less, and "long-term" otherwise. "Construction" indicates that the purchase transaction has associated documents indicating construction activity, excluding new constructions. "New construction" indicates that the purchase is identified as a transaction of a newly built home. "Holding period" stands for dummies based on holding-period sample deciles. "County FE" stands for county-level fixed-effect dummies. Standard errors are clustered over counties. *,**, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively.

Table A3
Relative performance after investor-specific exposure

Panel A: Return at purchase and sale date

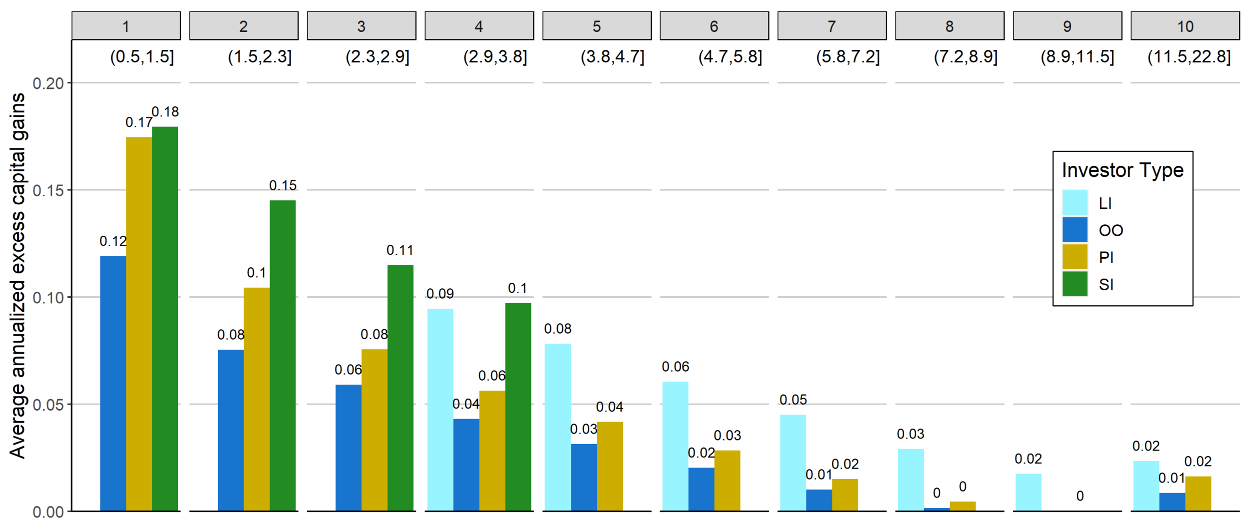
	PI		SI		LI		Adj. R-sq.	Observations
National index	0.013***	(0.001)	0.045***	(0.003)	0.039***	(0.002)	0.188	21,160,538
Stock market	0.013***	(0.001)	0.048***	(0.003)	0.031***	(0.001)	0.162	21,160,538
30-year mortgage	0.013***	(0.001)	0.049***	(0.003)	0.031***	(0.001)	0.163	21,160,538
County income	0.014***	(0.001)	0.048***	(0.003)	0.036***	(0.001)	0.200	20,492,659
County unemployment	0.013***	(0.001)	0.049***	(0.003)	0.035***	(0.001)	0.188	20,486,270
County density	0.011***	(0.002)	0.070***	(0.004)	0.032***	(0.002)	0.168	20,496,123
County turnover	0.013***	(0.001)	0.050***	(0.003)	0.030***	(0.001)	0.162	21,156,300
County LTV	0.012***	(0.001)	0.052***	(0.003)	0.028***	(0.001)	0.166	20,348,816

Panel B: Annualized changes over the holding period

	PI		SI		LI		Adj. R-sq.	Observations
National index	0.016***	(0.001)	0.076***	(0.004)	0.031***	(0.001)	0.332	20,967,087
Stock market	0.009***	(0.002)	0.049***	(0.003)	0.034***	(0.002)	0.171	20,967,094
30-year mortgage	0.016***	(0.001)	0.045***	(0.003)	0.046***	(0.002)	0.175	20,967,139
County income	0.007***	(0.002)	0.065***	(0.003)	0.015***	(0.003)	0.187	21,156,784
County unemployment	0.013***	(0.001)	0.051***	(0.003)	0.031***	(0.001)	0.230	21,170,761
County density	0.010***	(0.002)	0.071***	(0.004)	0.027***	(0.002)	0.173	21,178,517
County turnover	0.015***	(0.001)	0.056***	(0.003)	0.027***	(0.001)	0.193	21,178,869
County LTV	0.013***	(0.001)	0.052***	(0.003)	0.027***	(0.001)	0.172	20,444,399

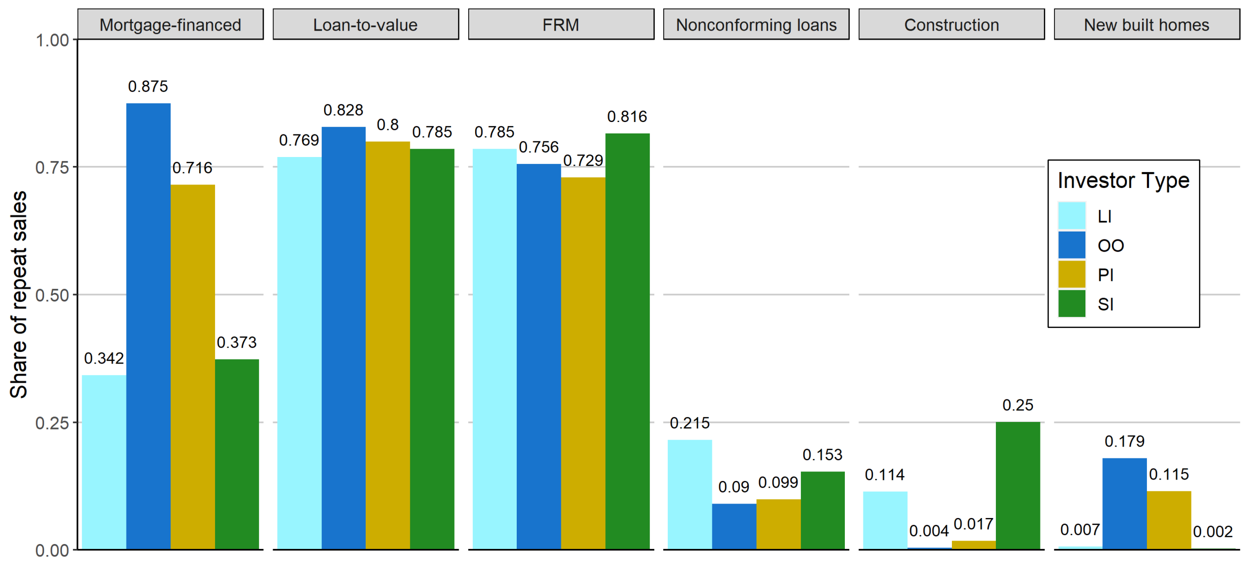
This table shows regression results with annualized excess capital gains as the dependent variable. The displayed coefficients are estimates for group-specific dummies. The reference category are owner-occupiers. For each regression, we interact the variable given in the first column with dummies for each buyer type, respectively. In each regression, the realization of the underlying variable at purchase and sale date is used, respectively. "National index" is the monthly S&P Case-Shiller US National Home Price Index, "Stock market" is the monthly Fama-French stock market factor, "30-year mortgage" is the monthly average 30-year mortgage interest rate, "County income" is the county's median income in a given year, "County unemployment" is the county's unemployment rate in the given year, "County density" is the county's population per square mile in a given year. "County turnover" is the quarterly number of transactions in a given period, and "County LTV" is the county's quarterly average mortgage amount divided by the average sales price (see Section 2.4). "PI", "SI", "LI" are dummy coefficients indicating that the corresponding trade was made by a private investor, short-term or long-term institutional investor, respectively. Private investors are defined as individuals who do not live in the underlying homes. A transaction is classified as "short-term" if the corresponding holding period is three years or less, and "long-term" otherwise. For Panel A, we interact each group with the given variable during the purchase and sale period, respectively. For Panel B, annualized changes over the given holding period are interacted with group dummies. The annualized variables are trimmed at the 0.5 and 99.5 percent level. All regressions include the same set of control variables: Dummies for construction, new construction, and holding-period sample deciles, as well as county-level fixed effects. Standard errors are clustered over counties. *,**, and *** indicate significance at the 5%, 1%, and 0.1% levels, respectively.

Figure A1
Capital gains by holding-period deciles



This figure plots average annualized excess capital gains realized by investor types for different holding-period deciles (given in years). Investor identity is assigned according to information associated with each individual trade: owner-occupiers (OO), i.e. households that live in the underlying home, private investors (PI) that are defined as private individuals who do not live in the traded home, short-term institutional investors (SI), i.e., trades made by institutional investors with a holding period of three or less years, and long-term institutional investors (LI), i.e., institutional sellers with a holding period of more than three years.

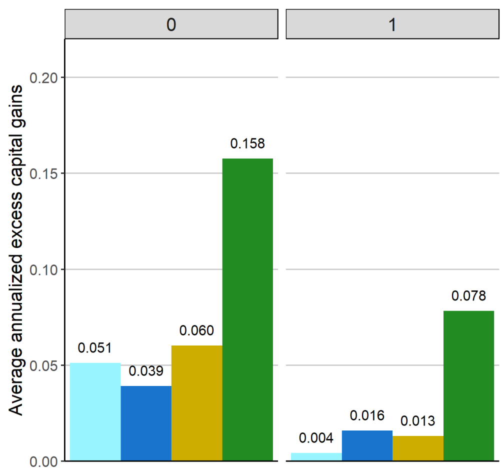
Figure A2
Purchase-specific characteristics by investor type



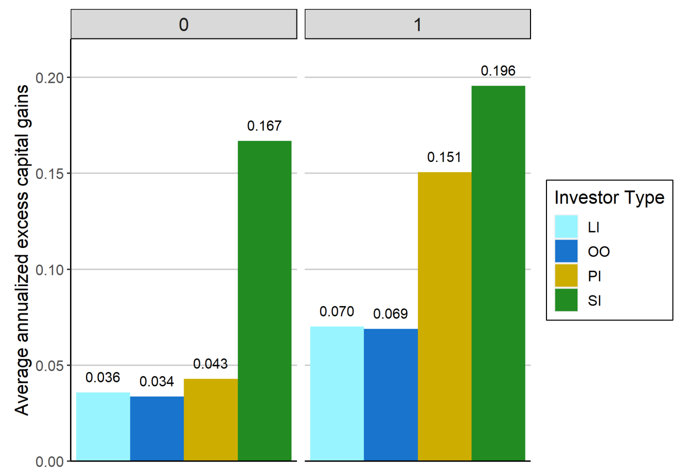
This figure plots the percentage share of repeat sales with specific purchase characteristics for different investor types. We compare the share of investor-specific mortgage-financed property investments, the average loan-to-value ratio (LTV), the share of fixed rate mortgages (FRM), as well as the share of non-conforming loans (not purchased by Fannie Mae or Freddie Mac for securitization) of debt-financed properties. We also document the percentage share of construction-related purchases and the share of newly built homes by investor types. Investor identity is assigned according to information associated with each individual trade: owner-occupiers (OO), i.e. households that live in the underlying home, private investors (PI) defined as private individuals who do not live in the traded home, short-term institutional investors (SI), i.e., trades made by institutional investors with a holding period of three or less years, and long-term institutional investors (LI), i.e., institutional sellers with a holding period of more than three years.

Figure A3
Investor performance by purchase characteristics

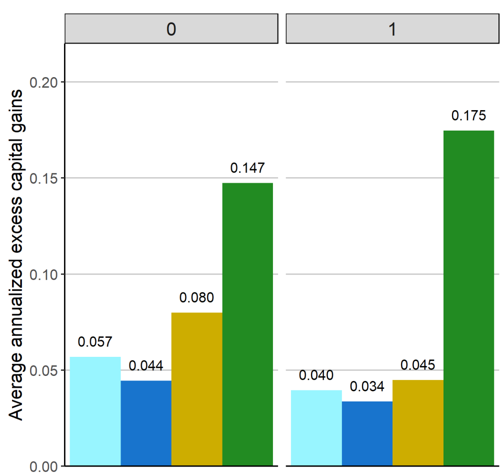
Panel A: New built homes



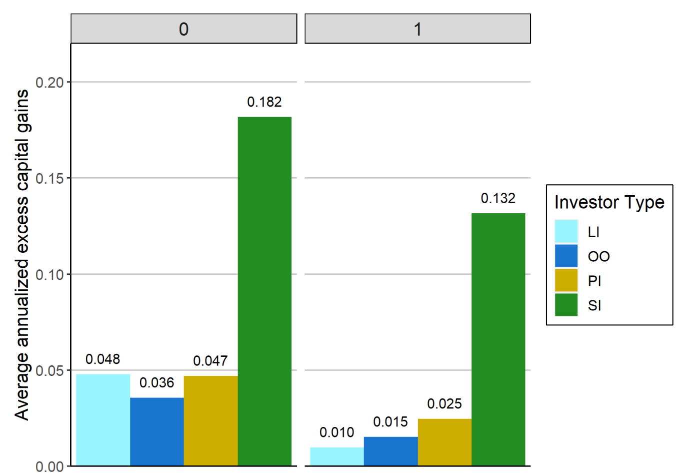
Panel B: Construction activity



Panel C: Mortgage-financed



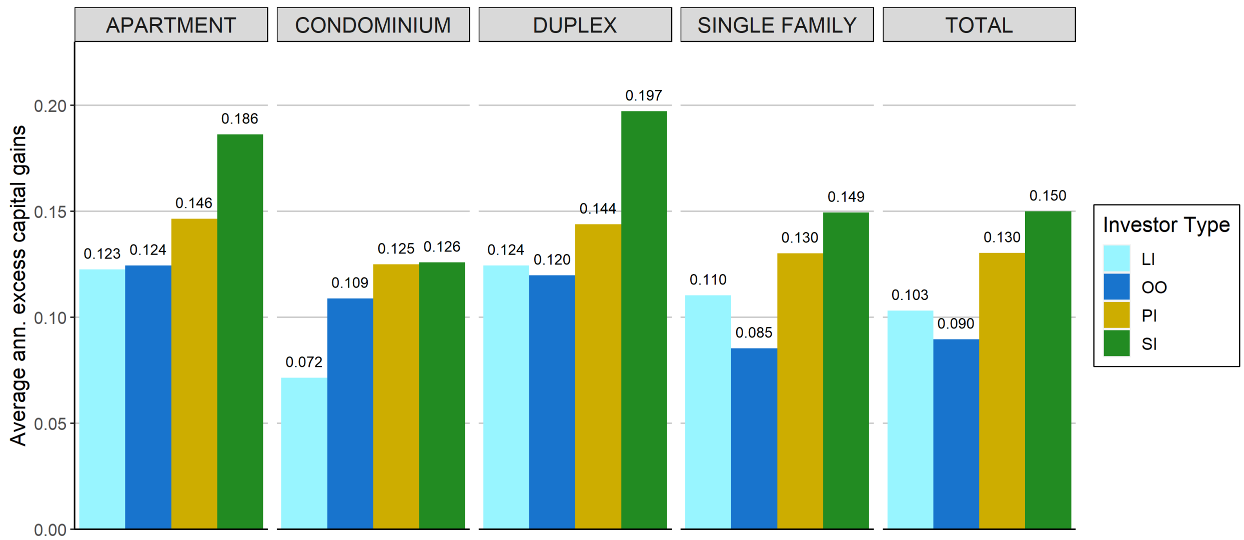
Panel D: Non-conforming loans



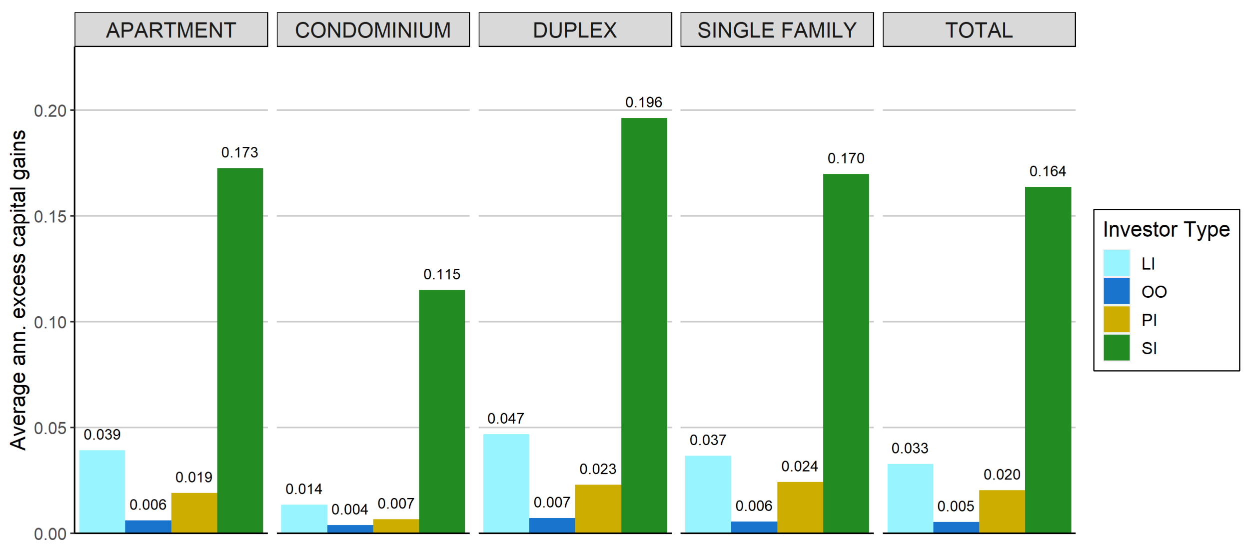
This figure plots average annualized excess capital gains realized by investor type separated by different purchase-characteristics, indicated by 1 if the characteristic is fulfilled for the purchase, and 0 else. Panels A to D compare performance-differences by investor types, conditional on investments in newly built homes, construction-related purchase activity, mortgage-financed investments and mortgage-related investments, which are either purchased by Fannie Mae and Freddie Mac for securitization purpose, or not. Investor identity is assigned according to information associated with each individual trade: owner-occupiers (OO), i.e. households that live in the underlying home, private investors (PI) defined as private individuals who do not live in the traded home, short-term institutional investors (SI), i.e., trades made by institutional investors with a holding period of three or less years, and long-term institutional investors (LI), i.e., institutional sellers with a holding period of more than three years.

Figure A4
Capital gains by housing sector and investor type

Panel A: Average capital gains realized between 1995 and 2006

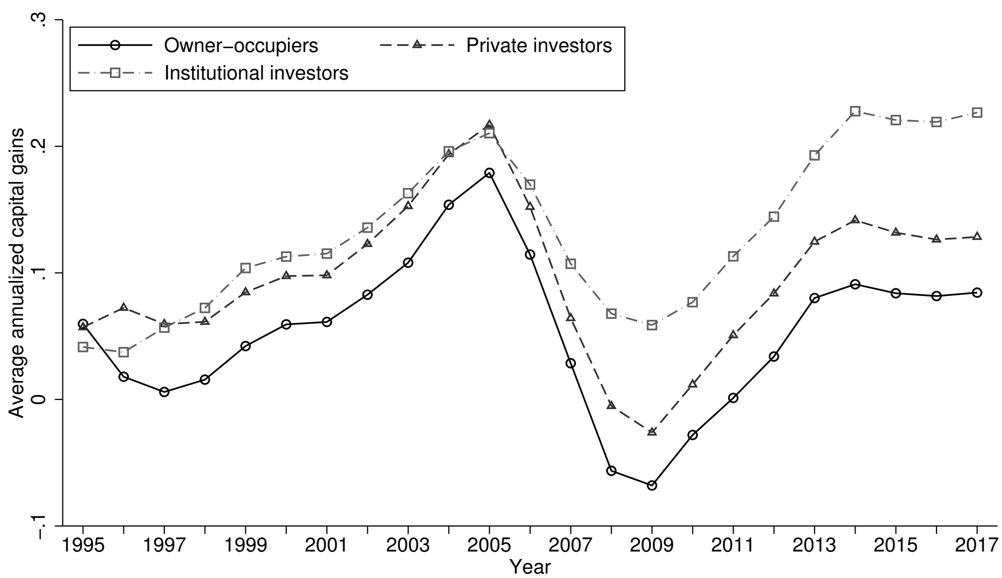


Panel B: Average capital gains realized between 2007 and 2017



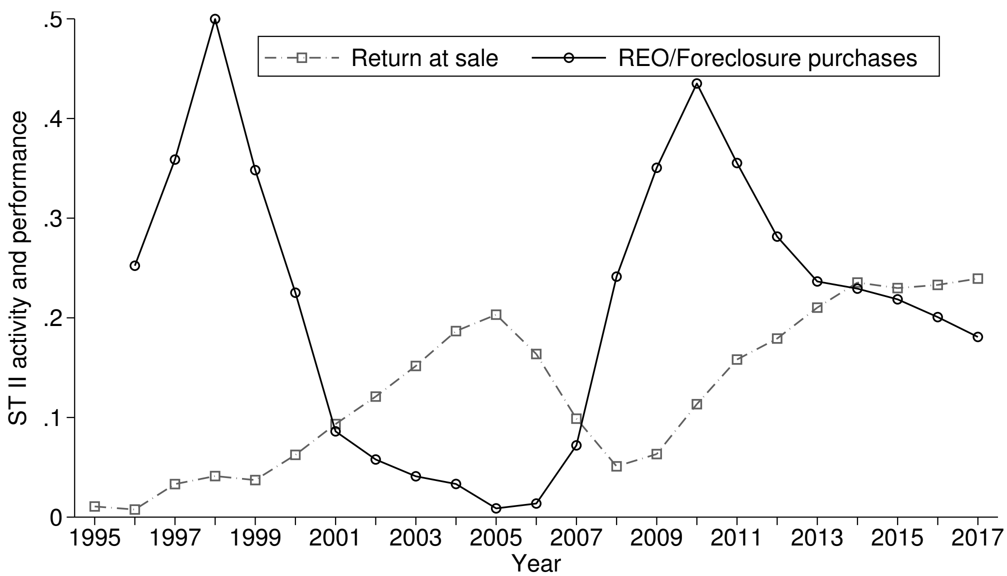
This figure plots average annualized excess capital gains realized by investor type separated by different subsectors (apartments, condominiums, duplex, single-family housing) and compared to the overall sample of repeat sales. Panel A depicts the average annualized excess capital gains realized (properties sold) between 1995 and 2006. Panel B shows the average annualized excess capital gains realized (properties sold) between 2007 and 2017. Investor identity is assigned according to information associated with each individual trade: owner-occupiers (OO), i.e. households that live in the underlying home, private investors (PI) defined as private individuals who do not live in the traded home, short-term institutional investors (SI), i.e., trades made by institutional investors with a holding period of three or less years, and long-term institutional investors (LI), i.e., institutional sellers with a holding period of more than three years.

Figure A5
Short-term capital gains in the US residential housing market



This figure shows short-term average annualized excess capital gains realized in a given year by a particular group of sellers, based on repeated sales of US residential homes with holding periods on three or less years traded between 1995 and 2017. Investor identities are uniquely assigned to each repeated sale: owner-occupiers (OO) are defined as households which live in the traded home, private investors (PI) are households which do not live in the traded home, institutional investors (SI) are companies selling within a holding period of three or less years.

Figure A6
Purchases of foreclosed properties and performance of SIs



The graph plots average annualized excess capital gains at time of sales (dashed-dotted, cubed line) and share of total short-term institutional investor (SI) REO/foreclosure *purchases* (relative to all observed SI purchases), depicted by the solid, circled line.

Swiss Finance Institute

Swiss Finance Institute (SFI) is the national center for fundamental research, doctoral training, knowledge exchange, and continuing education in the fields of banking and finance. SFI's mission is to grow knowledge capital for the Swiss financial marketplace. Created in 2006 as a public–private partnership, SFI is a common initiative of the Swiss finance industry, leading Swiss universities, and the Swiss Confederation.

swiss:finance:institute

c/o University of Geneva, Bd. Du Pont d'Arve 42, CH-1211 Geneva 4
T +41 22 379 84 71, rps@sfi.ch, www.sfi.ch

Impressum:
Arbeitskreis Quantitative Steuerlehre, arqus, e.V.
Vorstand: Prof. Dr. Ralf Maiterth (Vorsitzender),
Prof. Dr. Kay Blaufus, Prof. Dr. Dr. Andreas Löffler
Sitz des Vereins: Berlin

Herausgeber: Kay Blaufus, Jochen Hundsdoerfer,
Martin Jacob, Dirk Kiesewetter, Rolf J. König,
Lutz Kruschwitz, Andreas Löffler, Ralf Maiterth,
Heiko Müller, Jens Müller, Rainer Niemann,
Deborah Schanz, Sebastian Schanz, Caren Sureth-Sloane, Corinna Treisch

Kontaktadresse:
Prof. Dr. Dr. h.c. Dr. h.c. Caren Sureth-Sloane,
Universität Paderborn, Fakultät für
Wirtschaftswissenschaften,
Warburger Str. 100, 33098 Paderborn,
www.arqus.info, Email: info@arqus.info

ISSN 1861-8944