Residential Real Estate Traders:
Returns, Risk and Strategies

Marco Giacoletti* and Victor Westrupp†

Abstract

Asset dealers play a key role in providing liquidity to over-the-counter markets for both real and financial assets. We investigate the behavior and performance of asset dealers acting as middlemen in housing markets. We use a unique dataset covering house sales and remodeling jobs in the main urban areas of California from 1998 to 2012. On average, dealers have been able to extract substantial economic rents. However, dispersion in the performance of single transactions is large. Since most middlemen are involved in only a small amount of trades at a time, transaction-level risk is relevant. Despite positive average economic rents, asset dealers intermediate a smaller fraction of trades in housing markets than in other decentralized markets. The magnitude of transaction-level risk may help explain this fact.

*Stanford Graduate School of Business: Email mgiacol@stanford.edu

†Stanford Graduate School of Business: Email victor_westrupp@stanford.edu

1 Introduction

Residential housing markets are among the largest markets for real assets in the economy. Housing markets are illiquid and sales are negotiated over the counter. Many decentralized markets for financial and real assets are characterized by the presence of asset dealers. The business model of these intermediaries is based on superior trading technology or information, which is used to identify profitable transactions. The dealer buys an asset from a counter-party that has a low valuation and re-sells the same asset to a new counter-party who has a higher valuation. Thus, asset dealers are middlemen, able to extract value from the market, while improving asset allocation and providing liquidity. The degree to which asset dealers can extract value from the market is both a measure of the quality of their skills and of market inefficiency.

Relatively little is known about asset dealers in housing markets. Practictioners and business professionals monitor[1] the activity of these traders, which are frequently called "house flippers". However, it is apparent that the majority of house transactions take place directly in between households, with the assistance of brokers.

The fact that house flippers have not emerged as dominant players in housing markets is somewhat surprising. In this paper, we look for evidence of middlemen activity in residential real estate. Using data from the main urban areas of California (Los Angeles, Sacramento, San Diego and San Francisco), we show that the house flipping industry consists of a large number of small traders. We measure the economic rents extracted by these traders on a risk adjusted basis, using methodology from the private equity literature. On average, trades performed by middlemen deliver substantial and statistically significant abnormal compensation. We interpret this fact as evidence that, in the aggregate, house flipping activity generates value.

However, the dispersion of abnormal returns at the level of dealers' individual transactions is large. In fact, the bottom 25% of middlemen trades systematically under-performs the market. One hypothesis is that the large dispersion in outcomes is driven by differences in skill across traders. Even if a relevant fraction of individual transactions' can be explained by middlemen skill, the unexplained component remains substantial. In fact, unexplained transaction-level risk is so high that a one standard deviation shock would be large enough to entirely wipe out the average abnormal return. Risk at the level of individual transactions matters, since house flippers operate on a small scale: few of them take part in more than ten re-sales over a time-span of two years. We believe that the risk involved in individual transactions is potentially deterring a large number of small entrepreneurs from entering the house flipping market.

Our empirical study is based on a unique dataset created by merging micro-data on house sales from Corelogic[2] and a proprietary dataset on home remodeling permits provided by Buildzoom[3]. The remodeling permits data provide information on the timing and cost of major renovation jobs at the level of individual

[1]See for example the indexes reported at http://www.realtytrac.com/news/home-prices-and-sales/q2-2015-u-s-home-flipping-report/ or the commentary at http://www.trulia.com/blog/flipping-houses-real-estate-secrets-you-can-learn-from-house-flippers/.

[2]Licensed by Stanford University.

[3]We thank Issi Romen and Buildzoom for granting access to their database.

properties. Using these data we are able to keep track of large investments in house improvement taking place in between re-sales. The dataset covers house sales from the main urban areas of California (Los Angeles, Sacramento, San Diego and San Francisco) that were started over the years from 1998 to 2012.

Our goal is to identify trading activity by agents that professionally act as middlemen. Since middlemen activity consists of quickly matching sellers and buyers with different valuations, we focus our analysis on a "fast moving" segment of housing markets. This segment consists of houses that are re-sold after a holding period shorter than one year. For simplicity, we call these fast transactions "house flips". The "house flips" segment of the market is relatively small in our data, but not negligible: more than 30 billion dollars where invested in house flips (not even including remodeling expenses) across the four metropolitan areas in our study. The size of the house flipping segment changed over the recent housing cycles. It was larger during the housing boom and the recovery (respectively, from 2002 to 2005 and from 2010 to 2012) and the smallest during the housing bust (from 2006 to 2009).

A challenge faced when studying speculative behavior in housing markets is that there is no disclosure of trading motives in the data. Thus, the few existing papers in this area infer the motive of home buyers based on their past investment behavior. In particular, Haughwout et al. (2011) look for home owners who hold mortgages on multiple properties. Bayer et al. (2012) identify a home buyer as a "flipper" if she bought and resold at least two properties in the two years prior to the current purchase. We follow the methodology used in previous studies and define a home buyer as a professional middlemen if she bought and flipped at least two houses in the previous two years. With the word "flipped", we mean that the two houses were re-sold after an holding period shorter than a year.

Our analysis of middlemen performance is inspired from the literature in financial economics that studies the abnormal returns earned by professional money managers and private equity firms. In particular, each trade in our dataset involves multiple, irregularly spaced cash flows, and performance can be assessed only when a housing unit is sold. To address these issues, Kaplan and Schoar (2005) introduce the Public Market Equivalent (PME), which is a measure of the excess performance of a private equity fund with multiple investment flows with respect to a replicating portfolio invested in a public equity benchmark. We use the same methodology to calculate *Local Market Equivalents* (LMEs), which capture excess capital gains of real estate investment with respect to local ZIP code house price indexes (provided by Zillow).

Under the assumption that local ZIP code fluctuations have a loading of one on the capital gaisn of local housing units and that ZIP code price indexes span all relevant risk sources in the local market, the LME is a measure of abnormal performance for real estate investments. However, there is evidence that differences in house characteristics determine difference in capital gains (see for example Piazzesi et al. (2015)).

Thus, when analyzing LMEs we introduce controls for characteristics of the housing unit (number of bedrooms, number of rooms, house age, size in square feet) along with time and ZIP code fixed effects. It is important to note that abnormal performance is measured *ex-dividend* for both the house trade and the price index, since housing rents are excluded. We project our controls on LMEs separately for different metropolitan ares and different subsamples of vintages in the data, to account for the fact that relationships between characteristics and capital gains might change across different cities and different stages of the housing cycle.

House flips (sales with holding period shorter than one year) earn in several cities and vintage subperiods positive and statistically significant abnormal returns with respect to house sales with longer holding periods. This fact is consistent with evidence reported by Sagi (2015) in commercial real estate markets. Sagi (2015) shows, using a quantitative model, that this result can be generated by trading in illiquid markets where asset valuations are persistent but heterogeneous. Transactions after short holding periods only determined happen when low and high valuation agents meet.

Evidence of positive abnormal performance is even stronger when we focus on house flips executed by professional middlemen. For this groups of traders, house flips always produce positive average abnormal returns, across all metropolitan areas and subperiods. House flips performed by professional middlemen always outperform flips by agents that performed only zero or one flip in the previous two years. On average, flips executed by professional middlemen earn higher gains with respect to flips done by other traders (with the exception of the post housing crisis vintages in the city of Sacramento, where the difference in average performance is positive but not significant).

This is the first paper to assess the risk adjusted performance of middlemen in housing markets, taking into account local market fluctuations, house characteristics and remodeling expenses. The results from our analysis on average abnormal capital gains earned by middlemen are consistent previous findings by Bayer et al. (2012), who study the behavior of speculative traders in the housing market of Los Angeles. The authors find that middlemen are on average able to buy houses at a discount with respect to market prices, and to sell them back at or above market prices. However, Bayer et al. (2012) do not have data on remodeling activity, and do not study in detail the abnormal capital gains earned by middlemen.

Studying abnormal returns of house flips gives us insight on the performance of "successful" transactions carried out by middlemen. However, middlemen might not be able to flip all the properties that they acquire. Some of these properties might end up stuck in the middleman inventory, and get re-sold only after a substantial amount of time. We therefore extend our exercise be keeping track of all the properties that traders who flip houses buy over time. The average abnormal performance delivered with respect to houses re-sold over longer periods by non-flippers shrinks. However, the abnormal performance of professional mddilemen (who flipped at least two houses in the previous two years) remain positive and significant.

As mentioned above, a key fact concerning the industrial organization of the house flipping industry is that the operations run by middlemen are small in size. Less than 10% of owners engaging in "house flips" have flipped more than 10 houses in the previous two years. Thus, flippers might not just care about the average performance of transactions, but also about the individual risk involved in each single trade. We therefore focus our analysis on the professional middlemen, and use the output from our regression analysis to quantify the dispersion of individual trades performance, and we find that it is large. The inter-quartile ranges of abnormal capital gains are from three to five times larger than average abnormal capital gains.

As already explained above, a potential justification for the dispersion in transaction outcomes could be heterogeneity in middlemen skills. Following a similar approach to Korteweg and Sorensen (2016), we use an analysis of variance decomposition (ANOVA) to distinguish between variation in trade outcomes within and across different middlemen. Our findings suggest that there are significant differences in skill across middlemen: approximately one third of the variance in performances occurs across different middlemen. Nonetheless, the remaining dispersion is still extremely large. We interpret this finding as evidence that house flipping is a risky activity also for skilled middlemen.

The rest of the paper is organized as follows. Section 2 provides an overview of the structure and features of our dataset. Section 3 provides general facts on the "house flipping" segment of the housing markets included in our study. Section 4 discusses in detail how we define and measure abnormal performance for real estate transactions. Section 5 presents our empirical results on average abnormal performance. Section 6 studies individual transaction risk and performance persistence at the level of individual traders. Finally, section 7 concludes and presents the final remarks.

2 Data

We build a unique database consisting of micro-data on home re-sales and remodeling permits. We obtain micro-data on house sales from Corelogic[4], which collects information on house "deeds", consisting of regular sales, real estate ownerd (REO) sales and short sales. The "deeds" are collected from county records and provide details on transaction prices, as well as owner and seller names. Housing units are identified using Assessor Parcel Numbers (APNs, assigned to each plot of land by tax assessor of a specific jurisdiction for purposes of record-keeping) and addresses. Corelogic also provides information on house characteristics and geo-location based on tax assessment data.

Consistent with most of the literature on housing, we focus on "deeds" involving independent family houses, and consequently exclude condos from our dataset. While Corelogic data extend back to the the eighties and seventies, the quality of the records is higher for more recent entries and seems to be best starting from the late nineties. We therefore decide to use for our analysis the sample between January 1998 and December 2013. We then decide to focus on the main urban areas of California. This is both because these are important US housing market and because these markets have high quality coverage in the Corelogic database.

Our study focuses on the metropolitan areas of Los Angeles, Sacramento, San Diego and San Francisco. More precisely, we set the metropolitan area of Los Angeles to coincide with Los Angeles and Orange county. The metropolitan areas of San Diego and Sacramento correspond to their respective counties. Finally, for the metropolitan area of San Francisco, we select the Metropolitan Statistical Area (MSA) of San Francisco and Oakland (which includes Alameda county, Contra Costa county, Marin County, San Francisco County, San Mateo County), with the addition of Santa Clara County.

Data on remodeling contracts are provided by Buildzoom[5]. The company is an intermediary, matching homeowners to contractors for commercial and residential remodeling needs. To perform its business, Buildzoom has collected information on house remodeling permits from local census authorities. The

[4]More information on Corelogic is available at http://www.corelogic.com/industry/real-estate-solutions.aspx
[5]Information on Buildzoom is available at https://www.buildzoom.com/about.

data contain details on the kind of remodeling job that was performed (for example, kitchen renovation, plumbing, roof repairs and so on), the job cost and the fees paid to the local authority by the contractor.

We merge information from Corelogic and Buildzoom at the level of each housing unit. In the merge, we again identify each property based on its APN and address. More details on the data and the issues faced in the construction of the dataset can be found in appendix A. Finally, we build our dataset by keeping track of repeated sales of the same housing unit, and of remodeling investments in between re-sales.

We exclude from the final dataset all nominal sales, along with data on repeated sales with missing prices or geo-location information. We select the sample of repeated sales that were initiated between 1998 and 2012. A significant fraction of these sales are not followed by a re-sale within our data, and appear to have been held beyond December 2013. For these transactions, we compute a "synthetic" sale price as of December 2013, based on an hedonic regression on 2013 house prices. The synthetic price can be interpreted as a "fair" valuation of the house, as if it was sold at the end of sample. Details on the hedonic model can be found in appendix B.

Table 1 reports summary statistics of the data by year. The table shows the number of sales in each year and the median, mean and standard deviation of prices (in nominal terms). It also reports the fraction of housing units that were remodeled before a re-sale (or before the end of sample, for houses that were held till December 2013). This fraction is relatively high (always in the range between 12.5% and 16.5 %), a fact that highlights how it is important to keep track of remodeling activity when constructing data on house transactions. We also report for each year the fraction of houses that were held till the end of the sample, as well as the houses that were bought or sold through a distress sale. We define distress sales as either Real Estate Owned sales (REOs) or short sales.

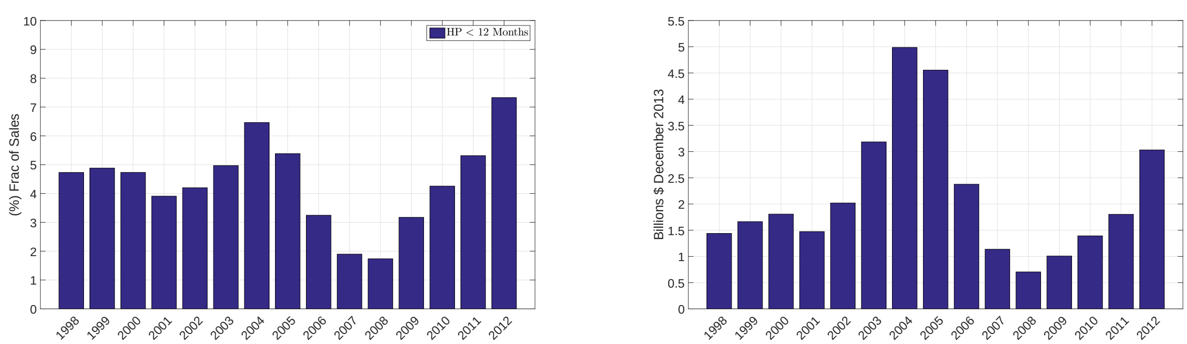
	N. Sales	Median Price	Mean Price	Std Price	Remodeled	No Re-Sale	Bought REO/Short	Sold REO/Short
1998	107,733	$211,500	$271,334	$223,267	14.49%	56.95%	4.59%	1.89%
1999	113,572	$226,000	$294,969	$268,983	15.13%	57.24%	1.71%	2.45%
2000	111,472	$250,000	$336,872	$337,465	16.20%	56.59%	2.11%	2.80%
2001	104,265	$275,000	$348,587	$296,647	16.37%	59.88%	2.17%	3.10%
2002	119,039	$329,000	$404,810	$329,422	16.49%	61.98%	1.52%	3.92%
2003	132,850	$390,000	$470,064	$344,981	16.47%	61.82%	0.62%	6.59%
2004	139,427	$475,000	$566,831	$443,480	16.47%	57.73%	0.40%	14.01%
2005	133,197	$557,000	$657,644	$451,704	14.90%	51.36%	0.26%	28.28%
2006	102,258	$593,000	$704,282	$507,165	14.33%	50.13%	0.80%	35.85%
2007	72,884	$610,000	$760,262	$679,332	15.71%	67.35%	9.42%	20.56%
2008	93,325	$385,000	$519,074	$550,896	15.33%	86.58%	52.83%	3.36%
2009	109,378	$335,000	$450,055	$479,485	14.56%	88.87%	52.77%	1.01%
2010	104,976	$361,000	$495,311	$554,619	13.73%	89.43%	44.83%	1.13%
2011	110,544	$349,000	$484,791	$557,483	13.27%	90.27%	44.98%	0.68%
2012	123,762	$380,000	$534,989	$614,545	12.42%	91.16%	35.84%	0.84%

Table 1: Summary statistics from the data. The statistics refer to the house sales that took place in each year. The first column contains the number of sales per year. Columns two, three and four respectively show mean price, median price and standard deviation of prices (in nominal terms). Column five contains the fraction of houses that were remodeled before re-sale (or before the end of sample), while column six reports the fraction of houses that were held till the end of sample. Finally columns seven and eight report the fraction of houses that were respectively bought and sold through distress sales.

3 Middlemen in Local Housing Markets

Houses are assets that are usually held by their owners for a long time. Evidence from US Census surveys in 2000 and 2010 suggests that the median holding period for a home is 8 years. However, there is a small but significant segment of the market which moves with a turnover shorter than one year. Panel (a) of figure 1 shows, out of all sales for each year from 1998 to 2012, the fraction of houses which were re-sold after less than 12 months. The data cover the metropolitan areas of Los Angeles, Sacramento, San Diego and San Francisco. The fraction of "fast" house trades is highly cyclical. It is close to 4% in the first part of the sample, between 1998 and 2002. It then raises to 6.5% in 2004, and drops to 2% in 2008 and 2009, at the bottom of the crisis. After the housing bust, the fraction of fast trades raises again, reaching a peak above 7% in 2012. In general, the submarket appears to be more active during booms and recoveries and shows a clear contraction during the housing bust.

Panel (b) of figure 1 shows, again for each year from 1998 to 2012, the volume of investments that went into house transactions with holding period shorter than one year. Values are expressed in terms of December 2013 dollars. Over the entire period and across the four metropolitan areas under analysis, the total investment is equal to 32 billion dollars. This calculation is done excluding any further remodeling expense taking place before re-sale. Investment volume is cyclical, similarly to what described for the fraction of fast sales out of all sales. In the first part of sample, the annual investment volume is in the order of magnitude of 1.5 billion $ per year. It then raises to 5 billion $ in 2004, and drops below 1 billion $ in 2009. Finally, it raises after the crisis, reaching 3 billion $ in 2012.



(a) (b)

Figure 1: Panel (a) shows the fraction of houses bought and resold with holding period smaller or equal than 12 months, out of all transactions collected for each year from 1998 to 2012. Panel (b) shows total investment volume in houses for transactions with holding period shorter than 12 months. The values in panel (b) are expressed in terms of December 2013 dollars.

An explanation for the existence of fast transactions in housing markets is that they are entirely driven by households that experienced an unexpected shock to housing demand or to their ability to finance homeownership, and therefore decided or where forced to quickly sell their home. An alternative is that there are traders transacting fast in housing markets and acting as asset dealers, or middlemen. We believe that empirical evidence points towards the second explanation. First, the one year horizon is too short to capture distress sales. Even if a homeowner was to start missing mortgage payments immediately after buying a house, the process leading to default and distress sale would take longer than a year. As a matter of fact, there are virtually no distress sales among the fast re-sales (in no year distress sales are more than 0.4% of total fast sales).

More crucially, a relevant fraction of the traders involved in fast sales have an history of dealing in this kind of transactions. In figure 2 we split fast transactions in each year based on how "active" the owner has been over the previous two years. From now on, we will call the fast transactions with holding period shorter than one year "flips". To build the figure, in each year we identify the owners involved in each house flip. We then count the number of houses that each owner has flipped in the previous two years across the four metropolitan areas in our study (Los Angeles, Sacramento, San Diego or San Francisco), excluding the current transaction. Then, for each year we can compute the fraction of flips carried out by owners with different degrees of "activity". We find that, on average across different years from 1998 to 2012, more than 40% of flips are carried out by owners who have flipped at least other two houses in the previous two years. This fraction reaches 60% if we focus on the post crisis period. Owners who have flipped multiple houses are involved in only approximately 1% of transactions for holding periods longer or equal than one year.

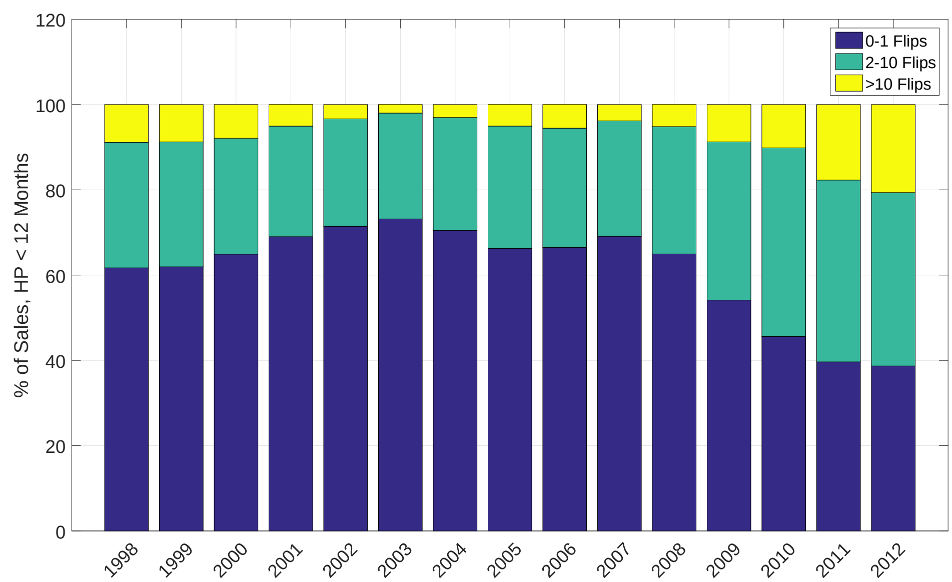


Figure 2: Fraction of house flips by owner's degree of activity in the local market. The statistics are reported year by year from 1998 to 2012.

Another important characteristic of the "house flips" segment is that a large fraction of transactions are undertaken by businesses or legal entities[6], rather than natural persons. The left hand-side panel of figure 3 shows that the fraction of houses bought and re-sold by legal entities is large when the holding

[6]Corelogic deed data include a flag that splits buyers in between corporations and individuals. To address potential weaknesses in the data, we also use text-search functions to select all buyers whose name contains corporate denominations, such as for example "INC","LLC", "& Co.", "Trust", "Investment", "Management" and other relevant key-words.

period is shorter than one year. Legal entities are involved in more than 60% of house flips in 2011 and 2012. This fraction is smaller in the earlier part of the sample, but always above 10%. For holding periods longer than one year, legal entities are always involved in less than 10% of transactions. We believe this is further evidence that house flips represent a separate segment of the housing market. This segment has a high presence of specialized investors, who use legal entities to separate the risks involved in their trading activities from their personal wealth.

Moreover, many of these legal entities are not simple real estate trusts. In figure 3 we dig a bit deeper onto the different kinds of legal entities that are flipping properties in the housing markets of California. For each year, we split the data based on the number of flips that the owners have undertaken over the previous two years, along the same lines of figure 2. We create three groups: zero or one flip, two to ten flips and more than ten flips. Figure 3 reports three panels, one for each group. Within the groups, we plot for each year the fraction of transactions undertaken by legal entities. We then that break down legale entities in between businesses with “INC” and “LLC” denomination and all other legal entities.

Transactions carried out by legal entities are present across all three groups, especially after 2007. The increase in the fraction of legal entities is clearly driven by the larger number of transactions undertaken by INCs and LLCs. Moreover, the fraction of legal entities is steeply increasing in the degree of activity of the owners.

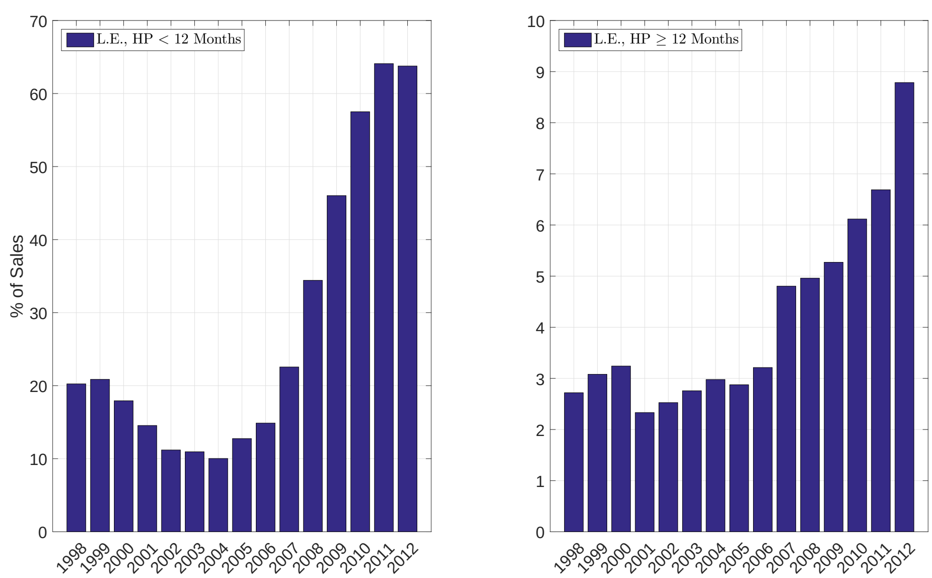


Figure 3: Fraction of transactions where the owner is a legal entity. The left hand-side panel refers to transactions with holding period shorter than one year, while the right hand-side panel refers to transaction with holding period greater or equal than one year. The statistics are reported year by year from 1998 to 2012.

Owners undertaking house flips also appear to mainly buy properties from and sell properties to non-flippers. We define non-flippers as owners who have not flipped any house in the previous two years. Table 2 focuses on the group of owners that flipped at least two houses in the previous two years. Panel

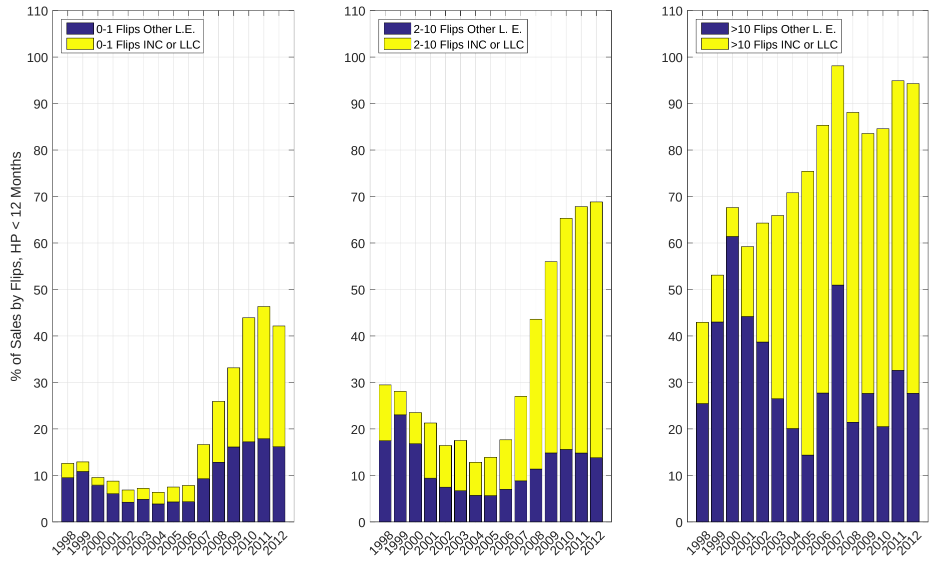


Figure 4: Fraction of legal entities (divided between INCs and LLCs and other entities) for different groups of owners. Holding periods shorter than one year. The left-hand side panel refers to owners that flipped zero or one house in the previous two years. The central panel refers to owners who flipped between two and ten houses in the previous two years and the right-hand size panel refers to owners who flipped more than 10 houses in the last two years. The statistics are reported year by year from 1998 to 2012.

(a) shows estimates of the fractions of houses that were bought from non-flippers, across the four different metropolitan areas and different time periods within our sample. Panel (b) repeats the same exercise for house sales. In both cases the fractions are very high, most of the time above 90%.

Table 3 shows the same estimates for all house flips (including the ones undertaken by owners that flipped zero or one house in the previous two years). Results are very similar to the ones in table 2. These findings further confirm that homeowners trading in this fast segment of the market resemble middlemen. These middlemen transfer assets in between “passive” homeowners, who trade infrequently.

	Panel (a): ≥ 2 Flips Group, Sold to non-Flipper				
	1998:2012	1998:2001	2002:2005	2006:2009	2010:2012
Los Angeles	76.92%	82.14%	67.73%	77.32%	79.45%
	(0.41%)	(0.94%)	(0.97%)	(1.10%)	(0.58%)
Sacramento	83.01%	82.61%	73.37%	87.66%	84.93%
	(0.84%)	(3.23%)	(2.22%)	(1.68%)	(1.08%)
San Francisco	78.74%	83.33%	63.41%	80.67%	81.23%
	(0.85%)	(3.40%)	(2.66%)	(2.56%)	(0.97%)
San Diego	83.33%	81.85%	77.34%	85.65%	86.00%
	(0.66%)	(2.21%)	(1.46%)	(1.61%)	(0.87%)
	Panel (b): ≥ 2 Flips Group, Bought from non-Flipper				
	1998:2012	1998:2001	2002:2005	2006:2009	2010:2012
Los Angeles	94.41%	92.90%	91.47%	96.19%	95.81%
	(0.23%)	(0.63%)	(0.58%)	(0.50%)	(0.29%)
Sacramento	95.58%	93.48%	92.46%	95.37%	97.06%
	(0.46%)	(2.10%)	(1.32%)	(1.07%)	(0.51%)
San Francisco	95.00%	90.00%	92.38%	97.90%	95.48%
	(0.45%)	(2.74%)	(1.47%)	(0.93%)	(0.52%)
San Diego	93.04%	89.77%	90.50%	94.94%	94.41%
	(0.45%)	(1.74%)	(1.02%)	(1.01%)	(0.58%)

Table 2: Fraction of houses sold to non-flippers, by metropolitan area and time period. Standard errors are in brackets.

	Panel (a): 0-1 Flips Group, Sold to non-Flipper				
	1998:2012	1998:2001	2002:2005	2006:2009	2010:2012
Los Angeles	74.45%	78.82%	69.69%	76.84%	78.44%
	(0.38%)	(0.82%)	(0.60%)	(0.91%)	(0.76%)
Sacramento	81.76%	82.62%	79.22%	85.66%	82.09%
	(0.78%)	(1.96%)	(1.30%)	(1.56%)	(1.58%)
San Francisco	76.04%	82.07%	68.36%	80.95%	79.23%
	(0.81%)	(1.84%)	(1.47%)	(1.87%)	(1.34%)
San Diego	80.99%	80.18%	77.81%	85.89%	83.31%
	(0.51%)	(1.21%)	(0.85%)	(1.07%)	(0.99%)
	Panel (a): 0-1 Flips Group, Bought from non-Flipper				
	1998:2012	1998:2001	2002:2005	2006:2009	2010:2012
Los Angeles	93.16%	91.96%	91.79%	95.12%	95.47%
	(0.22%)	(0.55%)	(0.36%)	(0.46%)	(0.38%)
Sacramento	93.28%	94.12%	90.23%	96.02%	95.44%
	(0.51%)	(1.22%)	(0.95%)	(0.87%)	(0.86%)
San Francisco	94.26%	91.49%	93.71%	96.83%	94.95%
	(0.44%)	(1.34%)	(0.77%)	(0.83%)	(0.73%)
San Diego	92.28%	88.90%	92.03%	94.57%	93.58%
	(0.35%)	(0.95%)	(0.55%)	(0.70%)	(0.65%)

Table 3: Fraction of houses sold to non-flippers, by metropolitan area and time period. Standard errors are in brackets.

In previous work, Bayer et al. (2012) show that middlemen in housing markets buy under-priced properties and re-sell them earning a premium. In doing this, they frequently engage in remodeling activities that improve house quality. This argument is also confirmed by studies conducted by pratictioners in the real estate industry[7]. While Bayer et al. (2012) develop an empirical strategy to identify houses that have been remodeled, they have no information on remodeling activity in their data. In our dataset, we are can keep track of major remodeling jobs (the ones requiring a permit) for each housing unit. Table reports the fraction of houses that underwent remodeling, for different categories of sales and across different metropolitan areas and time periods. Panel (a) shows remodeling frequencies for house re-sales across all holding periods. There are substantial differences across different metropolitan areas. Remodeling activity in our data is more than twice more frequent in San Francisco and Sacramento rather than in San Diego and Los Angeles. This result might be due to imperfect coverage of remodeling permits by Buildzoom. However, we believe it is reasonable to see large differences across the metropolitan areas in our study. In fact, the distribution of housing units' age varies widely across these areas: according to census data, 80% of housing units in the county of San Francisco where built before 1940, while the median age of housing units in Los Angeles is only 53 years.

Panels (b) and (c) report remodeling frequencies respectively for all houses with holding period shorter than one year and for all houses with holding period shorter than one year and traded by owners who have performed at least two other flips in the last two years. The interesting finding is that remodeling for these groups is in many cases as frequent or more frequent than for the sample including also transactions that had longer holding periods. This is particularly evident when focusing on the group of homeowners that have been flipping houses in the previous two years, as reported in panel (c). Thus, our findings suggest that remodeling plays an important role in house flips. Nonetheless, in our data most house flips do not involve remodeling, or at least do not involve a remodeling job that requires a permit.

4 Measuring the Performance of Middlemen

A key research question in this project is whether middlemen are able to extract value from local housing markets on a risk adjusted basis. Similar questions have been studied in the literature that asks whether active investors can extract value from financial markets. To analyze the risk adjusted performance of professional investors in liquid asset classes, researchers need first to find the risk factors for which investors demand risk compensation. They then need to estimate the loadings of these factors on the specific trading strategy pursued by an active investor (see for example Fama and French (2010)). The component of returns that is not explained by risk factors represents the abnormal or risk adjusted return generated by the active investor. This methodology has also been framed as a "horse race" of the investor against an alternative strategy, which is invested in portfolios that replicate exposures to the risk factors. However, as pointed out by Berk and van Binsbergen (2014), this interpretation is problematic when factors cannot be replicated using existing investment opportunities.

[7]See for example the material available at http://www.realtytrac.com/news/home-prices-and-sales/q2-2015-u-s-home-flipping-report/ or http://www.trulia.com/blog/flipping-houses-real-estate-secrets-you-can-learn-from-house-flippers/.

	Panel (a): Remodeling, All HP				
	1998:2012	1998:2001	2002:2005	2006:2009	2010:2012
Los Angeles	8.80%	10.15%	9.54%	8.75%	6.90%
	(0.03%)	(0.07%)	(0.05%)	(0.07%)	(0.08%)
Sacramento	19.99%	30.62%	24.21%	18.08%	13.72%
	(0.11%)	(0.30%)	(0.17%)	(0.21%)	(0.22%)
San Diego	11.38%	8.98%	12.24%	12.12%	10.61%
	(0.11%)	(0.15%)	(0.15%)	(0.23%)	(0.27%)
San Francisco	21.42%	22.37%	22.70%	21.14%	19.76%
	(0.07%)	(0.13%)	(0.11%)	(0.14%)	(0.16%)
	Panel (b): Remodeling, HP $<$ 12 months				
	1998:2012	1998:2001	2002:2005	2006:2009	2010:2012
Los Angeles	5.80%	3.90%	4.58%	5.90%	8.03%
	(0.15%)	(0.30%)	(0.23%)	(0.39%)	(0.31%)
Sacramento	16.87%	14.65%	10.66%	18.97%	21.49%
	(0.56%)	(1.56%)	(0.83%)	(1.31%)	(1.00%)
San Diego	6.60%	1.80%	4.74%	7.66%	8.36%
	(0.35%)	(0.56%)	(0.58%)	(1.02%)	(0.55%)
San Francisco	19.09%	7.90%	12.45%	21.86%	29.88%
	(0.41%)	(0.72%)	(0.58%)	(1.06%)	(0.83%)
	Panel (c): Remodeling, HP $<$ 12 months and Flips $\geq$ 2				
	1998:2012	1998:2001	2002:2005	2006:2009	2010:2012
Los Angeles	7.27%	4.93%	5.29%	7.84%	8.86%
	(0.25%)	(0.53%)	(0.46%)	(0.71%)	(0.41%)
Sacramento	21.66%	22.46%	14.07%	22.62%	23.99%
	(0.92%)	(3.55%)	(1.74%)	(2.12%)	(1.29%)
San Diego	7.65%	0.83%	6.10%	8.40%	8.36%
	(0.55%)	(0.83%)	(1.32%)	(1.80%)	(0.69%)
San Francisco	24.63%	16.17%	16.44%	21.73%	31.32%
	(0.76%)	(2.12%)	(1.29%)	(1.89%)	(1.16%)

Table 4: Fraction of houses remodeled, by metropolitan area and time period. Standard errors are in brackets.

The measurement of risk adjusted performance in housing markets carries many additional challenges. First, basic returns calculations are more complicated. Prices for individual real estate investments are observed only at the time a property is bought and re-sold. In addition, each transaction can involve multiple investment flows, due to remodeling expenses. Second, there is no information on dividends at the levels of individual houses. In other words, there are not micro-data on individual home rents or housing services. Third, it is challenging to come up with an appropriate risk model for a market that involves illiquid assets. Finally, since we observe returns only once for each transaction, the relationship between risk factors and returns cannot be estimated easily.

Our analysis focuses on house transactions undertaken by middlemen, who are trading residential properties fast. They are therefore unlikely to rent out their houses while looking for a buyer. Thus, we will for now focus on capital gains from house re-sales and leave rent rates outside of our analysis.

Figure 5 shows a diagram of the investment cashflows in housing unit i, bought at time t and sold at time $t+T$. The variable $P_{i,t}$ is the initial investment equal to the amount paid to buy the house. There can be H additional intermediate investments $D_{i,t+\tau_h}$ which capture remodeling costs, with $0 < \tau_h < T$. $P_{i,t+T}$ is the final payoff from selling the house.

A baseline benchmark against which we can compare the performance of a house flip is a price index for the local housing markets where the property is located. As local benchmarks, we download the series of Zillow Home Value Index (ZHVI) for ZIP codes in the California counties covered in our study [8]. Zillow

[8]The data are available at http://www.zillow.com/research/data/.

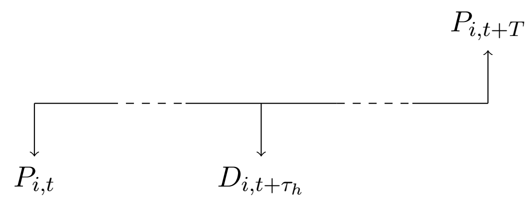


Figure 5: Cashflows diagram for a housing investment.

uses a statistical model to track the valuation of 100 million homes in the United States[9]. The ZIP code indexes are computed as the median values of homes in the area and are published monthly.

We measure abnormal capital gains of a house transaction over the corresponding ZIP code index using the public market equivalent (PME) methodology developed by Kaplan and Schoar (2005) for private equity funds. Performance measurement for private equity investments presents some of the same challenges faced in real estate. In particular, the funds have multiple investment flows and their value is observed only at the times of investment and liquidation. The PME measures the abnormal performance of a private equity fund with respect to a public equity benchmark.

We call our first measure of abnormal capital gains the *local market equivalent* (LME), since performance is defined in excess of a local house price index. The LME for property i from the example in figure 5 is equal to:

$$LME_{i,t} = \left(\frac{P_{i,t+T}}{R^{ZIP,i}_{t+T}} - \sum_{h=1}^{H} \frac{D_{i,t+\tau_h}}{R^{ZIP,i}_{t+\tau_h}} \right) \frac{1}{P_{i,t}} - 1 \tag{1}$$

Where $R^{ZIP,i}$ is the return on the price index for the ZIP code where property i is located. Notice that we omit rents from the performance calculation and house price indexes do not include rent payments. Thus, we are consistently comparing ex-dividend performance for the trade[10] and the benchmark index. Moreover, if we make the (admittedly strong) assumption that the median house tracked in the ZIP code index and property i make the same rent payments over the holding period, the LME is also a measure of total excess return with respect to the ZIP code index.

The LME is a satisfactory measure of abnormal capital gains only under the assumption that fluctuations in the local house price index span all the relevant risk sources in housing markets, and that the loading of the local index on individual house returns is always equal to one. However, it is reasonable to believe that houses with different characteristics have different price fluctuations. Moreover, the loading of the local house price index on individual properties might be different from one and might change over time. Following again the methodology of Kaplan and Schoar (2005), in our empirical analysis we try to deal with these issues by regressing LMEs onto house characteristics, as well as time and ZIP code fixed effects.

Note that the LME, both before and after controlling for additional information on house and trade characteristics, is a measure of the *gross* abnormal capital gain earned by a house transaction. When we

[9]More information on the methodology is available at http://cdn1.blog-media.zillowstatic.com/3/ZHVI-InfoSheet-04ed2b.pdf.

[10]While our data provide extensive micro-level information on transaction prices and remodeling costs, they do not include rent data at the level of individual properties or even individual ZIP codes covering the entire period of our study.

focus on professional investors, is likely that they will have financiers that demand a compensation for providing capital. We have limited and incomplete information on sources of financing for real estate traders in the data. Along the same lines, we do not know whether flippers are earning fees when performing their activity. Thus, we are not able to compute a measure of *net* abnormal returns, neither for the trader, nor for financiers.

Nonetheless, we do not believe this is necessarily a shortcoming of our work. As pointed out in Berk and Green (2004) and Berk and van Binsbergen (2014), the *net* abnormal return is not a measure of traders performance, since it is determined by the relative bargaining power of the trader and of those who provide capital. The fact that *gross* abnormal returns are greater than zero is a necessary condition for there being any opportunity to create value.

Berk and van Binsbergen (2014) also highlight that *gross* returns can be a misleading measure of the value extracted by professional traders from the market. Their argument, which is applied to the specific case of mutual funds, is that the professional investor determines her performance "in dollar terms" by selecting fund size. The authors argue that *gross* returns are a reliable measure of the rents extracted only when mutual funds are of very similar size. They introduce an alternative performance measure, called value added. Value added is equal to the gross abnormal return times investment size.

Variation in investment size is smaller across house investments than it is across mutual funds. Nonetheless, houses are lumpy investments and there is a fair amount of dispersion in house prices as can be seen in table 1. Thus, even if in our study we will mainly focus on estimating gross abnormal returns, we will also convert our results in value added terms.

5 Results on Abnormal Performance

5.1 Empirical Distribution of Local Market Equivalents

As the first step of our analysis, we compute LMEs for all house transactions in the data. Figure 6 reports the histograms for empirical distributions of LMEs. The figure is split in four panels, each one of them correspond to a different subsample of transactions, based on the year in which the transaction was initiated. The four blocks correspond respectively to the years from 1998 to 2001, from 2002 to 2005, from 2006 to 2009 and from 2010 to 2012. As explained in section 2, a fraction of our data consists of houses that were not re-sold before December 2013. For these transactions we build synthetic house prices as of December 2013 using an hedonic regression. Details of this procedure are available in appendix B.

We have previously assessed that fast house transactions (house flips) appear to constitute a separate subsegment of the market. We are now looking for evidence that LMEs in this subsegment have different properties with respect to the ones for houses held for longer holding periods. Thus, we plot separately in each panel of figure 6 the distribution of LMEs respectively for transactions with holding period equal or longer than one year, and holding period shorter than one year. The figure shows clearly that LMEs are larger for the shorter holding period, in each one of the four subperiods. Table 5 reports quantiles of LMEs distributions for the two subgroups in the different time periods. The median LME for transactions with holding period longer than one year is negative in all subperiods, and the ninetieth quantile of the LMEs distribution is in a range between 15% and 23% (depending on the subperiod).

The whole distribution of LMEs for house flips is shifted to the right when compared to the one for transactions with longer holding periods. The median LME for fast house trades is at its lowest value, slightly above 9%, in the period from 2002 to 2005. It is slightly smaller than 40% for the period from 2010 to 2012.

As discussed in section 4, LMEs computed with respect to local ZIP code indexes are only a first rough proxy for abnormal gains. There can be other factors driving average capital gains for individual housing units. Part of the wedge that we see in figure 6 and table 5 could be explained by the different composition of the groups of houses that are traded at short and long horizons. Our analysis in the next section will address this issue.

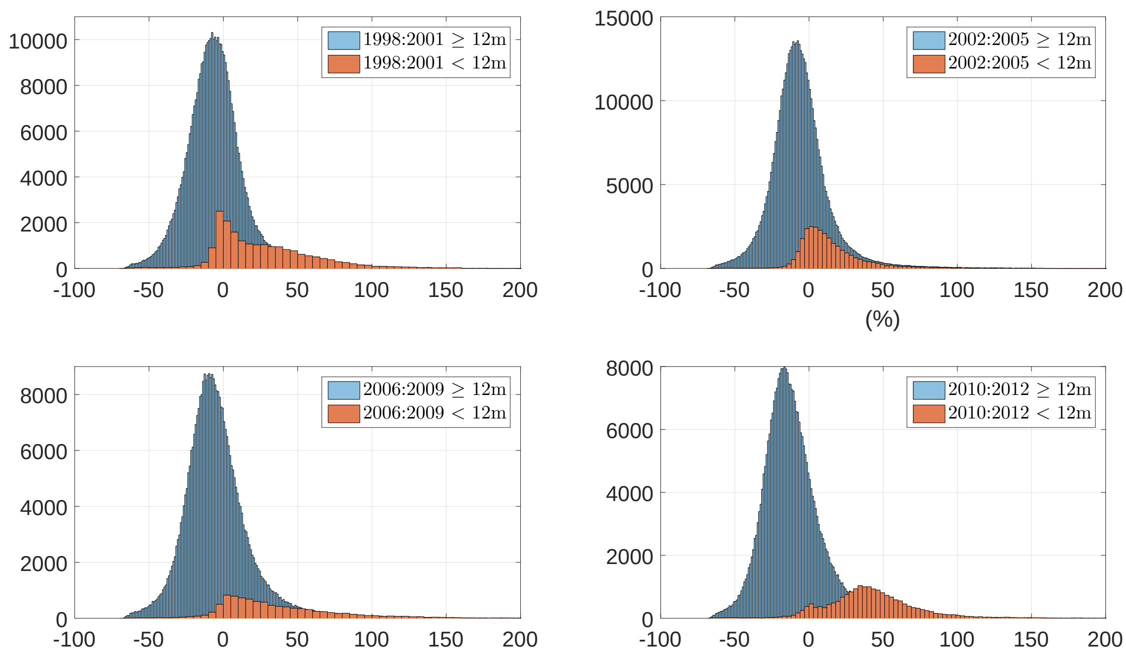


Figure 6: Distribution of LMEs for houses with holding period shorter and longer or equal than one year. Each panel corresponds to a different subperiod, based on when a house sale was started: from 1998 to 2001, from 2002 to 2005, from 2006 to 2009 and from 2010 to 2012.

In section 3 we introduced a distinction between different groups owners engaging in house flips, based on their past degree of activity in housing markets. In particular, we sorted owners in two groups. In the first we collect transactions by owners who engaged in at least two flips over the previous two years. In the second one we collect owners who flipped zero or one house in the previous two years. We believe the first group collects trades performed by investor that have more experience in the flipping market and are more likely to be professional traders. In figure 7 we compare the distribution of LMEs for the two groups of

	q10th	q25th	q50th	q75th	q90th
	Panel (a): 1998/1 to 2001/12				
	q10th	q25th	q50th	q75th	q90th
$HP < 12m$	-4.14%	1.94%	19.35%	45.25%	71.91%
$HP \geq 12m$	-27.69%	-17.09%	-6.06%	5.20%	18.74%
	Panel (b): 2002/1 to 2005/12				
	q10th	q25th	q50th	q75th	q90th
$HP < 12m$	-4.93%	0.70%	9.38%	22.36%	40.89%
$HP \geq 12m$	-26.95%	-17.38%	-7.59%	2.89%	15.55%
	Panel (c): 2006/1 to 2009/12				
	q10th	q25th	q50th	q75th	q90th
$HP < 12m$	-0.94%	8.28%	26.08%	54.85%	87.83%
$HP \geq 12m$	-27.71%	-17.54%	-6.52%	6.42%	23.16%
	Panel (d): 2010/1 to 2012/12				
	q10th	q25th	q50th	q75th	q90th
$HP < 12m$	4.85%	23.64%	39.49%	56.69%	77.06%
$HP \geq 12m$	-32.84%	-23.73%	-13.28%	-0.47%	15.98%

Table 5: Quantiles of the distribution of LMEs for houses with holding period shorter and longer or equal than one year.

traders and for house flips. The figure is again split in four panels, corresponding to transactions started from 1998 to 2001, from 2002 to 2005, from 2006 to 2009 and from 2010 to 2012. Across all subperiods, there is a clear wedge between the two distributions, in favor of the LMEs earned by more active traders. The wedge can be assessed even more clearly when looking at the quantiles of the distributions, which are reported in table 6. Of course, the concerns on composition effects expressed earlier in this section are still valid. We will address these concerns in the next section.

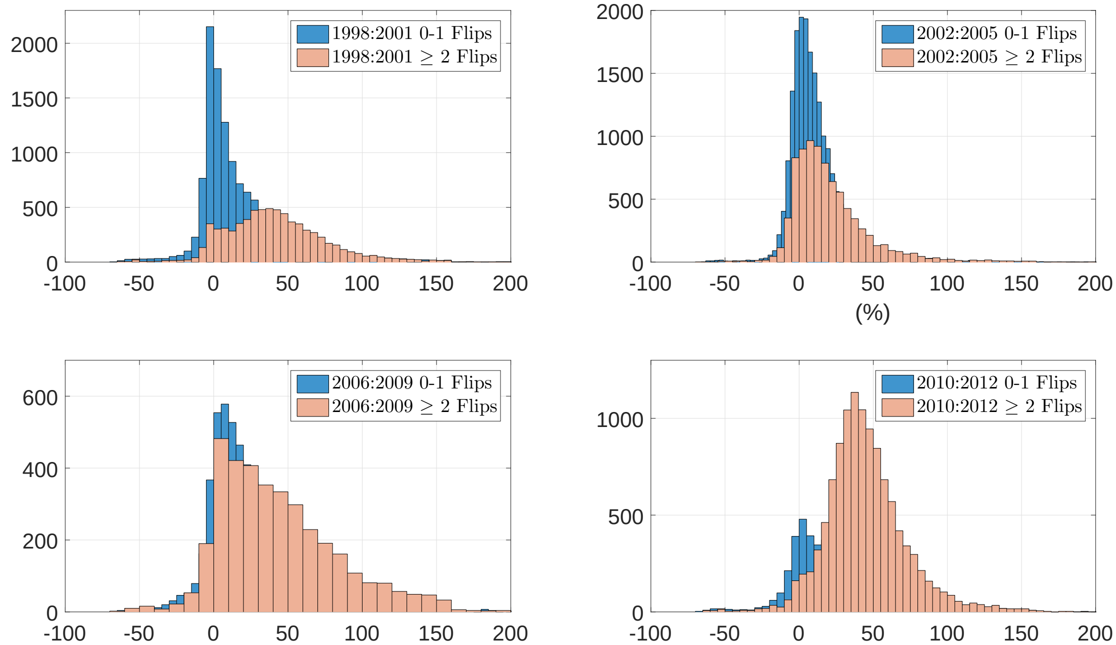


Figure 7: Distribution of LMEs for houses with holding period shorter than one year; owners who performed at least two flips in the last two years against owenrs that performed zero or one flip. Each panel corresponds to a different subperiod, based on when a house sale was started: from 1998 to 2001, from 2002 to 2005, from 2006 to 2009 and from 2010 to 2012.

	Panel (e): 1998/1 to 2001/12				
	q10th	q25th	q50th	q75th	q90th
0 - 1 Flips	-5.47%	-0.85%	9.05%	31.69%	60.19%
$\geq$ 2 Flips	1.30%	18.67%	38.69%	60.16%	83.79%
	Panel (f): 2002/1 to 2005/12				
	q10th	q25th	q50th	q75th	q90th
0 - 1 Flips	-5.51%	-0.17%	7.60%	18.99%	35.23%
$\geq$ 2 Flips	-3.29%	3.75%	14.75%	30.42%	50.85%
	Panel (g): 2006/1 to 2009/12				
	q10th	q25th	q50th	q75th	q90th
0 - 1 Flips	-2.26%	6.29%	21.24%	47.05%	79.52%
$\geq$ 2 Flips	1.25%	13.26%	36.20%	66.99%	100.73%
	Panel (h): 2010/1 to 2012/12				
	q10th	q25th	q50th	q75th	q90th
0 - 1 Flips	-1.33%	12.82%	34.46%	53.83%	75.30%
$\geq$ 2 Flips	15.85%	28.67%	42.11%	58.46%	77.73%

Table 6: Quantiles of the distribution of LMEs for houses with holding period shorter than one year; owners who performed at least two flips in the last two years against owenrs that performed zero or one flip.

5.2 Estimating the Abnormal Performance of House Flips

In this section, we provide estimates for the abnormal performance generated by house flips. So far we have measured abnormal capital gains as local market equivalents with respect to local ZIP code house price indexes. However, variation in house characteristics might drive differences in capital gains beyond what is captured by ZIP code fluctuations.

Thus, we set up our exercise so that we can compare the abnormal performance of a house flip against the abnormal performance of a transaction involving a similar house, but with a longer holding period. We are aware that we need to be careful with the interpretation of this comparison, since when we look at "longer holding periods" we are pooling different investment horizons. Nonetheless, we believe that the comparison can still reveal whether there is a substantial difference between the capital gains generated by the flipping market and the capital gains generated on average by the "slow moving" part of the housing market.

As a second step, we measure the abnormal performance of house flips performed by the more active traders (at least two flips in the last two years) with respect to the average abnormal performance of house flips undertaken by less active traders. Estimates of abnormal performance are obtained from regression equation 2:

$$LME_i = a_z + a_{yq,s} + a_{yq,e} + \mathcal{B}_{house} X_{i,house} + \beta_{12m} I_{i,HP<12m} + \beta_{12m,act} \left(I_{i,HP<12m} \times I_{i,act} \right) + v_i \tag{2}$$

Where LME_i is the LME for trade i. The coefficients a_z, $a_{yq,s}$ and $a_{yq,e}$ are respectively ZIP code fixed effects and year-quarter fixed effects for the starting and ending date of the trade. The vector $X_{i,house}$ contains controls for house characteristics. These are the log price at which the house was bought, the number of bedrooms, the number of bathrooms, the age of the structure, the square feet size of the house and a dummy in case the house was remodeled. We include the price at which the house was bought as a control variable since Piazzesi et al. (2015) and Piazzesi and Schneider (2016) show that the magnitude of house price fluctuations during the recent housing boom and bust have been different for more and less expensive houses and that this effect in not fully absorbed by ZIP code house price indexes or fixed effects. We include a dummy for remodeling to account for the fact that remodeling activity might on average generate value for homeowners.

The variable $I_{i,HP<12m}$ is a dummy, equal to one for house re-sales with holding period shorter than one year. $I_{i,act}$ is a dummy, equal to one for re-sales undertaken by active traders, who have flipped at least two houses in the previous two years. The coefficient β_{12m} captures the abnormal performance for house flips performed by non-active owners (who flipped zero or one houses over the previous two years). $\beta_{12m,act}$ captures the abnormal performance of house flips executed by active or experienced traders (who flipped at least two other houses over the previous two years) over flips by non-active traders. The sum of β_{12m} and $\beta_{12m,act}$ represents the abnormal performance of flips performed by experienced flippers with respect to house transactions with longer holding periods.

In this section, when estimating regression equation 2 from the data, we exclude from our sample all house re-sales that ended with a distress sale. We define as distress sales both real estate owned sales

(REOs) and short sales. This is because distress sales (almost) only affect holding periods longer than one year. Our aim is to find excess capital gains of house flips against comparable sales with longer holding periods. Including distress sales would give an unfair advantage to flips.

We estimate regression equation 2 separately for each metropolitan area. We first run the estimation over the entire sample of vintages. Due to the large fluctuations and changing conditions experienced by housing markets over the period under analysis, we re-estimate our regression on four subsambles for different vintages of house transactions. The subsamples correspond to vintages of house transactions that were started respectively from 1998 to 2001, from 2002 to 2005, from 2006 to 2009 and from 2010 to 2012. Parameter estimates for Los Angeles, Sacramento, San Diego and San Francisco are reported respectively in tables 7, 8, 9 and 10.

Figure 8 shows a summary of our estimates for abnormal performance across cities and subsamples of vintages. The figure reports abnormal performance for transactions performed respectively by non-active and active flippers, both computed against houses with longer holding periods. The figure also reports 95% confidence bands around our estimates. Results vary by city and subsample, but we consistently reject the null that abnormal capital gains delivered from houses flipped by experienced flippers are equal to zero, for all cities and subsamples. The abnormal performance is sizable, and is the highest for the vintage of trades initiated between 2006 and 2009. Evidence on the abnormal performance of flips performed by non-active flippers is more mixed. With the exception of San Francisco, after 2006 non-active flippers have statistically significant abnormal returns in all cities. However, their abnormal capital gains are approximately one half of the ones earned by their more experienced counter-parties.

Figure 9 summarizes estimates of the abnormal performance earned by house flips from active owners with respect to the ones from non-active owners. Again, results are reported for all cities and subsamples of vintages. The figure shows that abnormal performance is statistically significant for all cities and subsamples, with the exception of Sacramento for vintages starting during and after the housing crisis. In Los Angeles abnormal performance has a compelling "cyclical" pattern, it is equal to 10% before the boom and during the bust (vintages from 1998 to 2001 and from 2006 to 2009), and equal to 5% during the boom and the recovery (vintages form 2002 to 2005 and from 2010 to 2012).

We further expand our analysis in tables 7, 8, 9 and 10 in two directions. First, we test whether the abnormal performance that we find for active owners is entirely driven by transactions performed by businesses operating as established intermediaries. We include in our specification a dummy identifying flips that are carried out by active traders who also are business entities with denomination "INC" or "LLC". The regression equation becomes:

$$\begin{aligned} LME_i = a_z + a_{yq,s} + a_{yq,e} + \mathcal{B}_{house} X_{i,house} + \beta_{12m} I_{i,HP<12m} + \beta_{12m,act} \left(I_{i,HP<12m} \times I_{i,act} \right) \\ + \beta_{12m,act,buss} \left(I_{i,HP<12m} \times I_{i,act} \times I_{i,INC,LLC} \right) + v_i \end{aligned}$$

Where $I_{i,INC,LLC}$ is a dummy equal to one when the owner is an INC or LLC business. We find that INC and LLC on average businesses perform better than all other active traders (with a single exception for the metropolitan are of San Diego and the sample from 2002 to 2005). However, the difference between business and non-business active traders is not always significant. The coefficient $\beta_{12m,act}$ now identifies the abnormal performance of active traders that are not businesses. Its estimated values are smaller than the ones obtained from regression equation 2, but the difference is small and the coefficients remains significant. Thus, our conclusion is that the performance of active traders is not entirely explained by the activity of business entities.

With our second extension we want to test whether the evidence of abnormal performance is entirely determined by transactions where the flipper bought a house from a distress sale. Thus, we estimate the regression equation:

$$\begin{aligned} LME_i = a_z + a_{yq,s} + a_{yq,e} + \mathcal{B}_{house} X_{i,house} + \beta_{12m} I_{i,HP<12m} + \beta_{12m,act} \left(I_{i,HP<12m} \times I_{i,act} \right) \\ + \beta_{12m,act,buss} \left(I_{i,P<12m} \times I_{i,act} \times I_{i,INC,LLC} \right) + \beta_{distress} \left(I_{i,HP<12m} \times I_{i,distress} \right) v_i \end{aligned}$$

Where $I_{i,distress}$ is a dummy equal to one when the owner bought the house from an REO sale or a short sale. Our estimates reported in tables 7, 8, 9 and 10 show that distress sales explain an important fraction of the abnormal performance obtained by house flips from non-active traders, especially in Los Angeles, San Diego and San Francisco for the vintages from 2006 to 2009. However, controlling for houses bought from distress has hardly any impact on the estimates of abnormal performance for active traders with respect to non-active traders.

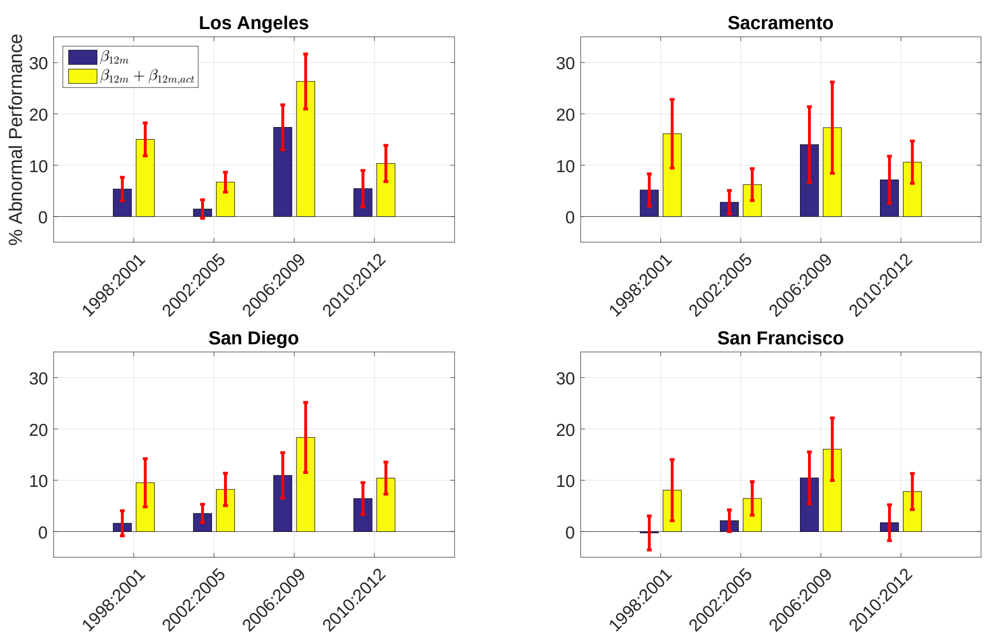


Figure 8: Abnormal performance for flips performed by active and non-active owners. Each panel corresponds to a different metropolitan area, results are reported separately for different blocks of vintages.

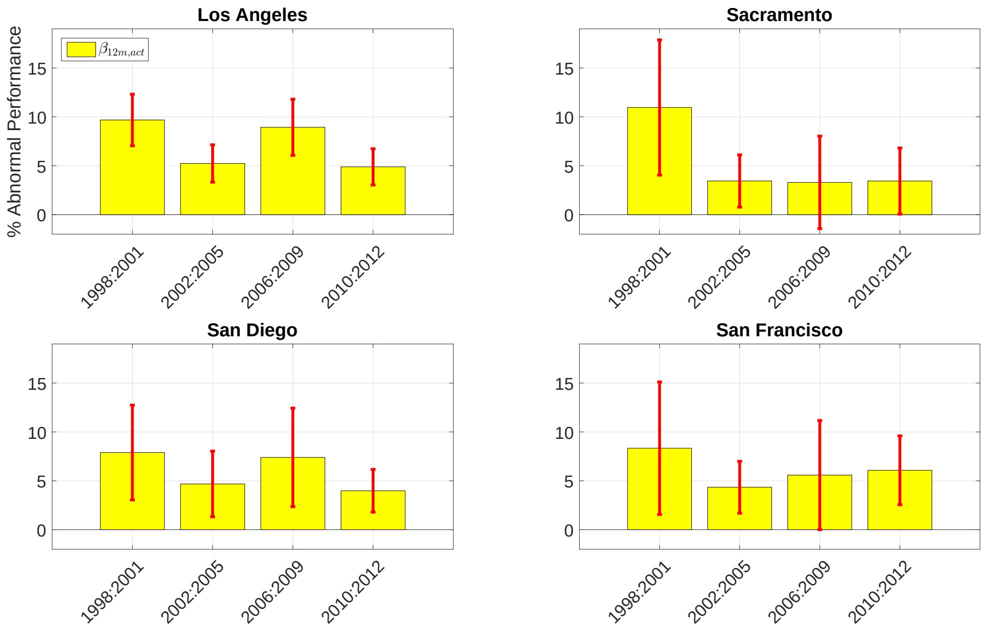


Figure 9: Abnormal performance for flips performed by active owners, with respect to flips performed by non-active owners. Each panel corresponds to a different metropolitan area, results are reported separately for different blocks of vintages.

	Metropolitan Area of Los Angeles														
	1998:2012			1998:2001			2002:2005			2006:2009			2010:2012		
$I_{HP<12m}$	0.0953 [10.1660]	0.0946 [9.8454]	0.0546 [7.1155]	0.0535 [4.5954]	0.0535 [4.5989]	0.0408 [3.6067]	0.0146 [1.5827]	0.0141 [1.5278]	0.0131 [1.4134]	0.1737 [7.7662]	0.1721 [7.6856]	0.0687 [3.1339]	0.0544 [3.0204]	0.0540 [2.9955]	0.0449 [2.3590]
$I_{HP<12m} \times I_{act}$	0.0986 [12.1966]	0.0670 [8.8377]	0.0465 [6.6844]	0.0968 [7.1862]	0.0954 [7.0366]	0.0772 [5.7977]	0.0523 [5.3764]	0.0343 [3.9152]	0.0322 [3.8049]	0.0894 [6.1031]	0.0600 [4.0168]	0.0449 [2.9959]	0.0489 [5.1605]	0.0249 [2.2571]	0.0234 [2.1479]
$I_{HP<12m} \times I_{act} \times I_{INC,LLC}$	– [–]	0.0879 [5.0693]	0.0606 [4.9578]	– [–]	0.0371 [0.8073]	0.0774 [1.6739]	– [–]	0.1323 [5.7218]	0.1369 [5.6931]	– [–]	0.0785 [3.1331]	0.0669 [2.8966]	– [–]	0.0395 [3.5808]	0.0397 [3.5993]
$I_{HP<12m} \times I_{distress}$	– [–]	– [–]	0.1413 [8.5338]	– [–]	– [–]	0.0916 [5.9711]	– [–]	– [–]	0.0824 [3.9331]	– [–]	– [–]	0.1637 [5.4055]	– [–]	– [–]	0.0129 [0.9478]
House-Specific Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Start Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
End Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R2	59.98 %	60.05 %	60.38 %	62.86 %	62.86 %	62.99 %	59.73 %	59.82 %	59.86 %	63.45 %	63.48 %	63.64 %	72.39 %	72.42 %	72.42 %
N. Obs.	439,697	439,697	439,697	81,253	81,253	81,253	132,016	132,016	132,016	111,725	111,725	111,725	114,703	114,703	114,703

Table 7: Regressions on local market equivalents. Standard errors clustered by Zip code and start year-quarter.

	Metropolitan Area of Sacramento														
	1998:2012			1998:2001			2002:2005			2006:2009			2010:2012		
$I_{HP<12m}$	0.0963	0.0963	0.0468	0.0517	0.0517	0.0376	0.0278	0.0278	0.0278	0.1401	0.1401	0.1071	0.0715	0.0715	0.0678
	[8.1236]	[8.0960]	[3.9819]	[3.2196]	[3.2193]	[2.2055]	[2.3754]	[2.3693]	[2.3701]	[3.7210]	[3.7241]	[2.5956]	[3.0521]	[3.0510]	[2.6701]
$I_{HP<12m} \times I_{act}$	0.0784	0.0684	0.0428	0.1096	0.1089	0.0860	0.0344	0.0308	0.0301	0.0330	0.0353	0.0339	0.0343	0.0326	0.0322
	[5.5098]	[5.1211]	[3.4844]	[3.1054]	[3.0496]	[2.4561]	[2.5221]	[2.0353]	[1.9793]	[1.3670]	[1.1240]	[1.0902]	[1.9924]	[1.7134]	[1.6365]
$I_{HP<12m} \times I_{act} \times I_{INC,LLC}$	–	0.0316	0.0157	–	0.0454	0.0862	–	0.0265	0.0278	–	−0.0069	−0.0083	–	0.0040	0.0042
	[–]	[1.3408]	[0.8086]	[–]	[1.0472]	[2.0740]	[–]	[0.8459]	[0.8781]	[–]	[−0.1227]	[−0.1511]	[–]	[0.2276]	[0.2355]
$I_{HP<12m} \times I_{distress}$	–	–	0.1358	–	–	0.0884	–	–	0.0602	–	–	0.0398	–	–	0.0045
	[–]	[–]	[5.5595]	[–]	[–]	[2.7369]	[–]	[–]	[0.8782]	[–]	[–]	[0.7913]	[–]	[–]	[0.1559]
House-Specific Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Start Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
End Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R2	68.57 %	68.58 %	68.81 %	63.85 %	63.85 %	63.98 %	61.59 %	61.59 %	61.60 %	75.15 %	75.15 %	75.16 %	78.84 %	78.84 %	78.84 %
N. Obs.	100,574	100,574	100,574	12,965	12,965	12,965	22,818	22,818	22,818	31,248	31,248	31,248	33,543	33,543	33,543

Table 8: Regressions on local market equivalents. Standard errors clustered by Zip code and start year-quarter.

	Metropolitan Area of San Diego														
	1998:2012			1998:2001			2002:2005			2006:2009			2010:2012		
$I_{HP<12m}$	0.1012	0.1013	0.0671	0.0162	0.0161	0.0125	0.0354	0.0353	0.0351	0.1094	0.1093	0.0412	0.0644	0.0641	0.0500
	[9.0321]	[8.9522]	[6.5036]	[1.3024]	[1.2942]	[0.9983]	[3.8940]	[3.8954]	[3.8248]	[4.8442]	[4.8276]	[1.3040]	[4.0654]	[4.0514]	[2.7865]
$I_{HP<12m} \times I_{act}$	0.0825	0.0602	0.0471	0.0789	0.0708	0.0635	0.0468	0.0483	0.0483	0.0739	0.0691	0.0618	0.0398	0.0223	0.0207
	[7.6982]	[6.0386]	[4.9980]	[3.1929]	[2.6108]	[2.5320]	[2.7327]	[3.0208]	[3.0304]	[2.8762]	[2.5391]	[2.3306]	[3.5696]	[1.7829]	[1.6575]
$I_{HP<12m} \times I_{act} \times I_{INC,LLC}$	–	0.0492	0.0331	–	0.0529	0.0668	–	−0.0149	−0.0147	–	0.0132	0.0066	–	0.0293	0.0291
	[–]	[3.9530]	[3.5242]	[–]	[1.1357]	[1.4957]	[–]	[−0.3458]	[−0.3407]	[–]	[0.3820]	[0.1971]	[–]	[2.7863]	[2.7962]
$I_{HP<12m} \times I_{distress}$	–	–	0.1023	–	–	0.1081	–	–	0.0386	–	–	0.1022	–	–	0.0211
	[–]	[–]	[5.7827]	[–]	[–]	[3.0032]	[–]	[–]	[0.7719]	[–]	[–]	[3.2774]	[–]	[–]	[1.6483]
House-Specific Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Start Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
End Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R2	56.39%	56.47%	56.98%	47.24%	47.25%	47.41%	38.23%	38.22%	38.22%	55.51%	55.51%	55.71%	77.15%	77.18%	77.20%
N. Obs.	41,079	41,079	41,079	7,220	7,220	7,220	13,211	13,211	13,211	10,201	10,201	10,201	10,447	10,447	10,447

Table 9: Regressions on local market equivalents. Standard errors clustered by Zip code and start year-quarter.

	Metropolitan Area of San Francisco														
	1998:2012			1998:2001			2002:2005			2006:2009			2010:2012		
$I_{HP<12m}$	0.0804	0.0804	0.0548	−0.0027	−0.0027	−0.0054	0.0211	0.0212	0.0209	0.1046	0.1041	0.0246	0.0173	0.0173	0.0128
	[6.8315]	[6.8002]	[5.0830]	[−0.1580]	[−0.1572]	[−0.3183]	[1.9613]	[1.9610]	[1.9293]	[4.0494]	[4.0495]	[1.0093]	[0.9702]	[0.9693]	[0.5566]
$I_{HP<12m} \times I_{act}$	0.0777	0.0601	0.0453	0.0834	0.0827	0.0745	0.0434	0.0368	0.0364	0.0559	0.0393	0.0346	0.0608	0.0541	0.0530
	[7.1430]	[5.1329]	[3.7597]	[2.4115]	[2.3261]	[2.1416]	[3.2110]	[2.8094]	[2.7741]	[1.9627]	[1.3402]	[1.2094]	[3.3799]	[2.5842]	[2.4109]
$I_{HP<12m} \times I_{act} \times I_{INC,LLC}$	–	0.0650	0.0498	–	0.0058	0.0207	–	0.0568	0.0575	–	0.0827	0.0776	–	0.0160	0.0164
	[–]	[2.3748]	[1.9376]	[–]	[0.0768]	[0.2749]	[–]	[1.2275]	[1.2425]	[–]	[0.9276]	[0.8891]	[–]	[0.6592]	[0.6728]
$I_{HP<12m} \times I_{distress}$	–	–	0.0981	–	–	0.1326	–	–	0.0479	–	–	0.1348	–	–	0.0068
	[–]	[–]	[4.6465]	[–]	[–]	[2.8144]	[–]	[–]	[1.1088]	[–]	[–]	[3.1651]	[–]	[–]	[0.3052]
House Char	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Start Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
End Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R2	58.28 %	58.29 %	58.36 %	64.76 %	64.76 %	64.79 %	63.69 %	63.70 %	63.70 %	67.45 %	67.46 %	67.51 %	72.50 %	72.50 %	72.50 %
N. Obs.	270,496	270,496	270,496	39,174	39,174	39,174	78,939	78,939	78,939	72,287	72,287	72,287	80,096	80,096	80,096

Table 10: Regressions on local market equivalents. Standard errors clustered by Zip code and start year-quarter.

5.3 Estimating the Abnormal Performance of House Flippers

In the previous section, we focused on measuring the abnormal performance of *house flips*. This is not the same as measuring the performance of *house flippers*. In fact, a owner who operates on the fast moving segment of the housing market might be able to successfully flip only a fraction of the properties that she is trading. The remaining part of her portfolio would consist of unsuccessful flips, that would eventually be liquidated on a longer holding period. Thus, the overall performance of the flipper has to be measured considering all the properties that she is trading.

In this section we focus on estimating the average performance for a trade undertaken by a flipper, no matter whether that trade ended up being a flip. We can then rewrite our regression equation as:

$$LME_i = a_z + a_{yq,s} + a_{yq,e} + \mathcal{B}_{house} X_{i,house} + \beta^{tot}_{12m} \tilde{I}_{i,HP<12m} + \beta^{tot}_{12m,act} \left(\tilde{I}_{i,HP<12m} \times \tilde{I}_{i,act} \right) + v^{tot}_i \quad (3)$$

Fixed effects and house-specific controls are the same as described in the previous section. The variable $\tilde{I}_{i,HP<12m}$ is dummy, now equal to one for all house re-sales undertaken by owners who have flipped at least one house. The variable $\tilde{I}_{i,act}$ is a dummy equal to one for all re-sales undertaken by owners who have flipped at least two houses in the previous two years. The coefficients β^{tot}_{12m} and $\beta^{tot}_{12m,act}$ have similar interpretations to β_{12m} and $\beta_{12m,act}$ in the previous section. We again run the regression separately for the four different metropolitan areas in our dataset and for four blocks of vintages: from 1998 to 2001, from 2002 to 2005, from 2006 to 2009 and from 2010 to 2012.

Results are reported in tables 11, 12, 13 and 14. Estimates of the abnormal performance of active and non-active house flippers are shown in figure 10. We again find that the abnormal performance of active traders is positive and statistically significant across all metropolitan areas and vintage subsamples. However, the magnitude of abnormal returns is smaller. When considering all transactions (not only flips) the abnormal performance of non-active traders deteriorates substantially. On the other hand, it is interesting that the estimates of the excess performance of active traders with respect to non-active traders do not change with respect to what we observed in the previous section. Figure 11 compares estimated values of $\beta_{12m,act}$ and $\beta^{tot}_{12m,act}$ across all metropolitan areas and time periods. We can see that the values of the coefficients are extremely close, especially if we consider the magnitude of the confidence intervals around point estimates.

Along the same lines as in section 5.2, we extend our regression setting to test whether the abnormal performance of active traders is entirely driven by business entities and whether it is just determined by transactions that involved buying houses through distress sales. Similarly to what we found in the previous section, none of these effects fully explains the performance earned by active traders.

	Metropolitan Area of Los Angeles														
	1998:2012			1998:2001			2002:2005			2006:2009			2010:2012		
$\tilde{I}_{HP<12m}$	0.0355	0.0356	0.0132	0.0407	0.0407	0.0323	−0.0043	−0.0043	−0.0050	0.0978	0.0970	0.0281	0.0079	0.0080	−0.0048
	[7.0938]	[6.9530]	[3.0735]	[5.0402]	[5.0422]	[4.1220]	[−0.6223]	[−0.6285]	[−0.7248]	[8.0115]	[7.9764]	[2.7102]	[0.6726]	[0.6756]	[−0.4643]
$\tilde{I}_{HP<12m} \times I_{act}$	0.1138	0.0843	0.0460	0.0956	0.0942	0.0737	0.0531	0.0362	0.0341	0.0978	0.0719	0.0356	0.0417	0.0214	0.0105
	[14.9266]	[12.8184]	[7.7424]	[7.4456]	[7.2853]	[5.6722]	[6.3725]	[4.8452]	[4.6158]	[6.1583]	[4.6368]	[2.4576]	[5.2805]	[2.2461]	[1.1288]
$\tilde{I}_{HP<12m} \times I_{act} \times I_{INC,LLC}$	–	0.0851	0.0467	–	0.0431	0.0848	–	0.1332	0.1376	–	0.0679	0.0520	–	0.0334	0.0336
	[–]	[4.5600]	[3.6190]	[–]	[0.9556]	[1.8765]	[–]	[6.0208]	[6.0092]	[–]	[3.1004]	[2.6074]	[–]	[2.8798]	[2.9962]
$\tilde{I}_{HP<12m} \times I_{distress}$	–	–	0.1881	–	–	0.0991	–	–	0.0872	–	–	0.2104	–	–	0.0490
	[–]	[–]	[12.5129]	[–]	[–]	[6.8182]	[–]	[–]	[4.1757]	[–]	[–]	[7.0768]	[–]	[–]	[3.8245]
House-Specific Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Start Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
End Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R2	59.18 %	59.26 %	59.97 %	62.75 %	62.75 %	62.90 %	59.41 %	59.50 %	59.55 %	63.08 %	63.11 %	63.54 %	72.26 %	72.28 %	72.33 %
N. Obs.	441,014	441,014	441,014	81,293	81,293	81,293	132,380	132,380	132,380	111,856	111,856	111,856	114,740	114,740	1147,40

Table 11: Regressions on local market equivalents. Standard errors clustered by Zip code and start year-quarter.

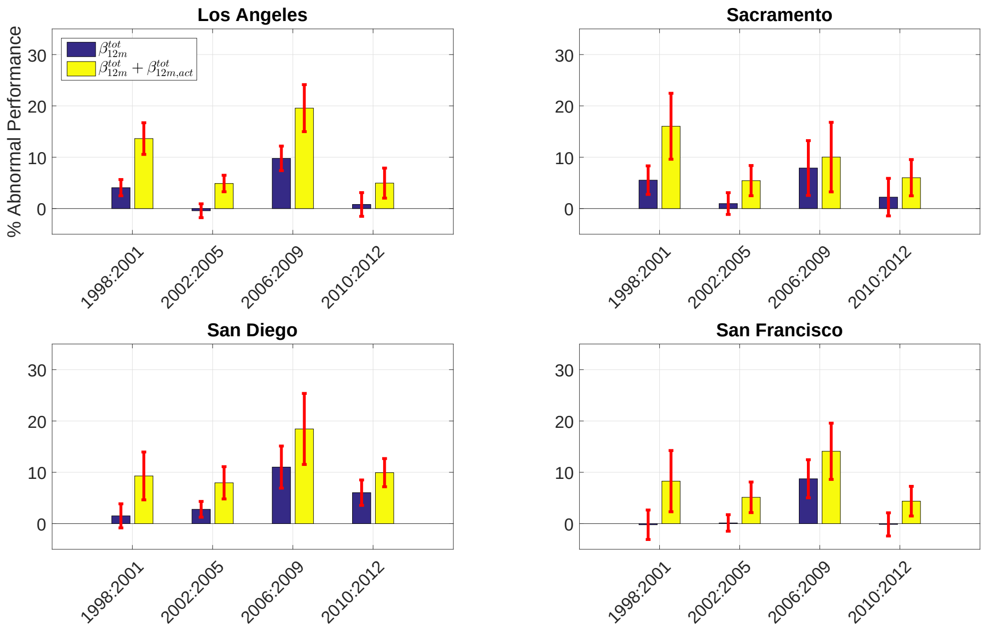


Figure 10: Abnormal performance for all trades performed by active and non-active owners. Each panel corresponds to a different metropolitan area, results are reported separately for different blocks of vintages.

Figure 11: Abnormal performance respectively for all trades and for flips only performed by active owners. Performance is computed in excess of capital gains to flips carried out by non-active owners. Each panel corresponds to a different metropolitan area, results are reported separately for different blocks of vintages.

	Metropolitan Area of Sacramento														
	1998:2012			1998:2001			2002:2005			2006:2009			2010:2012		
$\tilde{I}_{HP<12m}$	0.0590	0.0590	0.0231	0.0553	0.0556	0.0440	0.0097	0.0097	0.0097	0.0790	0.0791	0.0257	0.0222	0.0223	0.0022
	[6.8535]	[6.8509]	[3.1131]	[3.8996]	[3.9217]	[2.9782]	[0.8970]	[0.8938]	[0.8974]	[2.8976]	[2.9009]	[1.1825]	[1.1918]	[1.1951]	[0.1204]
$\tilde{I}_{HP<12m} \times I_{act}$	0.0842	0.0818	0.0375	0.1050	0.1081	0.0850	0.0447	0.0423	0.0415	0.0213	0.0290	0.0166	0.0379	0.0407	0.0252
	[6.7113]	[6.4073]	[3.2238]	[3.0673]	[3.0230]	[2.4029]	[3.1082]	[2.7582]	[2.7174]	[0.9965]	[1.0693]	[0.6254]	[2.4640]	[2.3878]	[1.4035]
$\tilde{I}_{HP<12m} \times I_{act} \times I_{INC,LLC}$	–	0.0074	−0.0062	–	−0.1081	−0.0713	–	0.0173	0.0186	–	−0.0228	−0.0259	–	−0.0063	−0.0034
	[–]	[0.3623]	[−0.3695]	[–]	[−1.0049]	[−0.6204]	[–]	[0.5825]	[0.6197]	[–]	[−0.5461]	[−0.6411]	[–]	[−0.3708]	[−0.2002]
$\tilde{I}_{HP<12m} \times I_{distress}$	–	–	0.1716	–	–	0.0851	–	–	0.0612	–	–	0.1273	–	–	0.0630
	[–]	[–]	[8.3733]	[–]	[–]	[2.6282]	[–]	[–]	[0.8952]	[–]	[–]	[2.8374]	[–]	[–]	[2.7398]
House-Specific Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Start Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
End Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R2	68.10%	68.10%	68.57%	63.89%	63.89%	64.01%	61.00%	61.00%	61.01%	74.87%	74.88%	74.99%	78.76%	78.76%	78.81%
N. Obs.	100,709	100,709	100,709	12,965	12,965	12,965	22,869	22,869	22,869	31,265	31,265	31,,265	33,560	33,560	33,560

Table 12: Regressions on local market equivalents. Standard errors clustered by Zip code and start year-quarter.

	Metropolitan Area of San Diego														
	1998:2012			1998:2001			2002:2005			2006:2009			2010:2012		
$\tilde{I}_{HP<12m}$	0.0915	0.0916	0.0611	0.0152	0.0151	0.0117	0.0279	0.0279	0.0276	0.1102	0.1101	0.0527	0.0603	0.0602	0.0465
	[8.9542]	[8.8885]	[6.5566]	[1.2678]	[1.2618]	[0.9829]	[3.5548]	[3.5548]	[3.4960]	[5.2654]	[5.2478]	[1.9409]	[4.7803]	[4.7768]	[3.4369]
$\tilde{I}_{HP<12m} \times I_{act}$	0.0882	0.0665	0.0478	0.0778	0.0694	0.0620	0.0517	0.0532	0.0532	0.0743	0.0707	0.0597	0.0390	0.0218	0.0184
	[8.0839]	[6.5566]	[5.0053]	[3.2641]	[2.6578]	[2.5842]	[3.0770]	[3.3785]	[3.3908]	[2.9925]	[2.6518]	[2.2736]	[3.5467]	[1.7432]	[1.4695]
$\tilde{I}_{HP<12m} \times I_{act} \times I_{INC,LLC}$	–	0.0480	0.0308	–	0.0545	0.0684	–	−0.0148	−0.0146	–	0.0098	0.0043	–	0.0287	0.0284
	[–]	[3.7996]	[3.2405]	[–]	[1.1822]	[1.5497]	[–]	[−0.3433]	[−0.3377]	[–]	[0.2760]	[0.1275]	[–]	[2.7290]	[2.7422]
$\tilde{I}_{HP<12m} \times I_{distress}$	–	–	0.1102	–	–	0.1087	–	–	0.0413	–	–	0.0956	–	–	0.0260
	[–]	[–]	[6.2523]	[–]	[–]	[3.0097]	[–]	[–]	[0.8257]	[–]	[–]	[3.2666]	[–]	[–]	[2.0901]
House-Specific Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Start Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
End Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R2	56.16%	56.23%	56.86%	47.23%	47.24%	47.41%	38.01%	38.01%	38.01%	55.52%	55.52%	55.73%	77.08%	77.12%	77.16%
N. Obs.	41,144	41,144	41,144	7,222	7,222	7,222	13,234	13,234	13,234	10,206	10,206	10,206	10,457	10,457	10,457

Table 13: Regressions on local market equivalents. Standard errors clustered by Zip code and start year-quarter.

	Metropolitan Area of San Francisco														
	1998:2012			1998:2001			2002:2005			2006:2009			2010:2012		
$\tilde{I}_{HP<12m}$	0.0519	0.0519	0.0323	−0.0022	−0.0022	−0.0045	0.0013	0.0013	0.0011	0.0874	0.0872	0.0422	−0.0015	−0.0015	−0.0108
	[6.2978]	[6.2823]	[4.6797]	[−0.1507]	[−0.1514]	[−0.3074]	[0.1612]	[0.1622]	[0.1314]	[4.6152]	[4.6269]	[2.1742]	[−0.1309]	[−0.1305]	[−0.9015]
$\tilde{I}_{HP<12m} \times I_{act}$	0.0801	0.0696	0.0441	0.0850	0.0859	0.0778	0.0500	0.0458	0.0454	0.0536	0.0460	0.0319	0.0453	0.0442	0.0371
	[7.5863]	[6.1152]	[3.6524]	[2.4586]	[2.4242]	[2.2319]	[3.8024]	[3.6415]	[3.5940]	[2.1991]	[1.7980]	[1.2308]	[2.8181]	[2.3115]	[1.7940]
$\tilde{I}_{HP<12m} \times I_{act} \times I_{INC,LLC}$	–	0.0384	0.0208	–	−0.0079	0.0063	–	0.0333	0.0341	–	0.0367	0.0326	–	0.0025	0.0040
	[–]	[1.4111]	[0.8201]	[–]	[−0.1096]	[0.0885]	[–]	[0.7851]	[0.8031]	[–]	[0.4069]	[0.3774]	[–]	[0.1076]	[0.1679]
$\tilde{I}_{HP<12m} \times I_{distress}$	–	–	0.1283	–	–	0.1311	–	–	0.0573	–	–	0.1280	–	–	0.0309
	[–]	[–]	[6.4037]	[–]	[–]	[2.7579]	[–]	[–]	[1.3475]	[–]	[–]	[2.8907]	[–]	[–]	[1.7974]
House-Specific Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Start Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
End Year-Quart FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R2	58.07 %	58.07 %	58.22 %	64.76 %	64.76 %	64.80 %	63.37 %	63.37 %	63.38 %	67.42 %	67.42 %	67.49 %	72.42 %	72.42 %	72.44 %
N. Obs.	270,835	270,835	270,835	39,175	39,175	39,175	79,066	79,066	79,066	72,330	72,330	72,330	80,107	80,107	80,107

Table 14: Regressions on local market equivalents. Standard errors clustered by Zip code and start year-quarter.

6 Risk, Value Added and Persistence of Performance

In this section we focus more closely on the economic value of flipping activity from the perspective of individual house flippers. In the previous sections we have shown that owners who have been actively flipping houses earn on average positive and statistically significant performance on their transactions. However, our finding does not immediately imply that engaging in house flipping activity is economically sensible from the perspective of homeowners. A first issue is that we measure gross abnormal gains. The value extracted from the market by flippers has to be shared with the principals who are providing them capital. Depending on the bargaining power of the two sides[11], the share left as compensation to flippers could be a small fraction of the gross abnormal capital gains. Second, and most crucially, most house flippers in our data only transact a limited number of houses. Figure 2 shows that, for each vintage of house re-sales, less than 10% of house flips are undertaken by owners who have flipped more than ten homes in the previous two years.

It is possible that some flippers may operate in many other local housing markets, outside the metropolitan areas covered in our study. However, as far as our data are concerned, the scale of flippers operations is usually quite small. In particular, it is not large enough to wipe away the effects of the idiosyncratic risk involved in each individual transaction. Moreover, anecdotal evidence[12] suggest that many flippers at least start off as small entrepreneurs, who have limited capacity of dealing with multiple properties. Thus, the idiosyncratic risk of each transaction might play an even greater role from the perspective of a new player entering the market.

We can get a rough measure of idiosyncratic risk for the transactions undertaken by active house flippers using results from regression equation 3. In fact, we can compute an estimate of the abnormal capital gain for each individual transaction executed by non-active flippers as:

$$\widehat{LME}_{j,a} = \hat{\beta}^{tot}_{12m,act} + \hat{v}^{tot}_{j}$$

Where j is any trade undertaken by the owners that were selected in the data as active flippers (they performed at least two flips in the previous two years). Table 15 reports estimates of the mean (which is by construction equal to $\hat{\beta}^{tot}_{12m,act}$), inter-quartile range and top and bottom quartile of the distribution of $\widehat{LME}_{j,a}$. The statistics are reported for the different metropolitan areas and for the blocks of vintages already used in the previous two sections. We can see that the dispersion of $\widehat{LME}_{j,a}$ is large. The ratio of the inter-quartile range over the mean performance is in between three and five for Los Angeles. The situation is similar for San Diego, while the ratio is even larger for San Francisco and Sacramento. The bottom quartile of trades earns a negative performance across all cities and vintages.

[11]We attended information sessions held by FortuneBuilders. It appears that many house flippers are not getting financing from banks, but rather from non-bank Private Money Lenders (PMLs). One of these institutions in the area of San Jose (in Santa Clara county) is Grand Coast Capital, which offers debt financing with a 75% loan to value ratio for loan sizes between \$ 100,000 and \$ 3,000,000. Information on loan programs can be found at www.grandcoastcapital.com . Grand Coast Capital seems to be also willing to provide equity-like financing under "specific circumstances".

[12]There are several companies providing training to individuals or households interested in becoming real estate flippers. An example is FortuneBuilders, which provides information on its services at http://www.thanmerrill.com/

	Metropolitan Area of Los Angeles			
	Mean	q75th - q25th	q75th	q25th
1998:2001	9.56%	25.32%	21.33%	−3.99%
2002:2005	5.31%	20.16%	14.51%	−5.65%
2006:2009	9.78%	32.49%	24.78%	−7.71%
2010:2012	4.17%	25.52%	16.43%	−9.09%
	Metropolitan Area of Sacramento			
	Mean	q75th - q25th	q75th	q25th
1998:2001	10.50%	30.02%	25.03%	−4.99%
2002:2005	4.47%	17.68%	12.76%	−4.92%
2006:2009	2.13%	35.63%	20.07%	−15.55%
2010:2012	3.79%	28.93%	18.65%	−10.29%
	Metropolitan Area of San Diego			
	Mean	q75th - q25th	q75th	q25th
1998:2001	7.78%	18.20%	16.62%	−1.59%
2002:2005	5.17%	15.37%	12.04%	−3.33%
2006:2009	7.43%	27.46%	21.57%	−5.88%
2010:2012	3.90%	18.61%	13.39%	−5.22%
	Metropolitan Area of San Francisco			
	Mean	q75th - q25th	q75th	q25th
1998:2001	8.50%	28.58%	20.97%	−7.61%
2002:2005	5.00%	20.87%	14.82%	−6.06%
2006:2009	5.36%	36.72%	21.14%	−15.58%
2010:2012	4.53%	30.15%	19.19%	−10.96%

Table 15: Distribution of abnormal capital gains for house re-sales executed by active flippers. Abnormal performance is measured with respect to re-sales undertaken by non-active flippers.

Focusing on the distribution of abnormal performance with respect to non-active flippers might appear as a particularly harsh benchmark. We can also estimate abnormal performance with respect to transactions undertaken by non-flippers. Those can be computed using again equation 3 as:

$$\widehat{LME}_{j,b} = \hat{\beta}^{tot}_{12m} + \hat{\beta}^{tot}_{12m,act} + \hat{v}^{tot}_{j}$$

Table 6 reports mean performance (equal to $hat\beta^{tot}_{12m} + \hat{\beta}^{tot}_{12m,act}$) and inter-quartile ranges across the different metropolitan areas and vintages. Results are not substantially different with respect to what already discussed for table 15. The ratio of the inter-quartile range over the mean performance is large, and the bottom quartile of transactions is earning a negative performance in almost all metropolitan ares and vintages.

We have so far assessed the performance of house re-sales in terms of gross abnormal capital gains. However, as suggested in the mutual funds literature by Berk and van Binsbergen (2014), active investors endogenously select the size of their investments. In this sense, we can think that professional investors target value creation in terms of dollar value added, rather than percentage performance. A measure of the dollar value added by each transaction is:

$$\widehat{VA}_j = \widehat{LME}_j \times P_j$$

	Metropolitan Area of Los Angeles			
	Mean	q75th - q25th	q75th	q25th
1998:2001	13.64%	25.32%	25.40%	0.08%
2002:2005	4.88%	20.16%	14.08%	-6.08%
2006:2009	19.57%	32.49%	34.56%	2.07%
2010:2012	4.96%	25.52%	17.23%	-8.29%
	Metropolitan Area of Sacramento			
	Mean	q75th - q25th	q75th	q25th
1998:2001	16.03%	30.02%	30.56%	0.54%
2002:2005	5.44%	17.68%	13.73%	-3.95%
2006:2009	10.03%	35.63%	27.97%	-7.66%
2010:2012	6.01%	28.93%	20.87%	-8.06%
	Metropolitan Area of San Diego			
	Mean	q75th - q25th	q75th	q25th
1998:2001	9.29%	18.20%	18.13%	-0.07%
2002:2005	7.96%	15.37%	14.83%	-0.54%
2006:2009	18.45%	27.46%	32.59%	5.13%
2010:2012	9.93%	18.61%	19.42%	0.81%
	Metropolitan Area of San Francisco			
	Mean	q75th - q25th	q75th	q25th
1998:2001	8.28%	28.58%	20.75%	-7.83%
2002:2005	5.13%	20.87%	14.95%	-5.92%
2006:2009	14.10%	36.72%	29.88%	-6.84%
2010:2012	4.38%	30.15%	19.04%	-11.11%

Table 16: Distribution of abnormal capital gains for house re-sales executed by active flippers. Abnormal performance is measured with respect to re-sales undertaken by non-flippers.

Where P_j is the price at which property j was bought. $\widehat{LME}_j$ can be set equal to either $\widehat{LME}_{j,a}$ or $\widehat{LME}_{j,b}$, depending on whether we are interested in the value added with respect to non-active flippers, or in the value added with respect to non-flippers.

Table 17 reports statistics for value added with respect to non-active flippers. The table shows mean value added and inter-quartile ranges for transactions taking place across the four different metropolitan areas in our study and different vintages. Roughly speaking, average value added per trade was in a range between 10,000 and 25,000 dollars in the pre-crisis period. Value added drops substantially across all areas in the last vintage block, which consists of the years from 2010 to 2012.

Table 18 reports the statistics for value added with respect to non-flippers transactions. Average value added is substantially larger, especially for the sample of vintages between 2006 and 2009, where abnormal performance is larger than 50,000 $ per transaction in Los Angeles and San Diego, and higher than 42,000 $ per transaction in San Francisco. We can again see a substantial drop in value added for vintages from 2010 to 2012. Table shows that the average gross abnormal capital gains for this last block of vintages are similar to the ones in the 2002-2005 period. Moreover, real house prices in the 2010-2012 period are comparable to the ones in the early part of the 2002-2005 sample. Thus, we believe that the drop in value added reveals that active flippers have targeted cheaper homes in more recent years.

	Metropolitan Area of Los Angeles			
	Mean	q75th-q25th	q75th	q25th
1998:2001	$12,748$	$31,074$	$25,649$	$-5,425$
2002:2005	$14,918$	$72,891$	$48,823$	$-24,068$
2006:2009	$15,731$	$78,716$	$57,602$	$-21,114$
2010:2012	$2,248$	$53,549$	$31,929$	$-21,620$ $

	Metropolitan Area of Sacramento			
	Mean	q75th-q25th	q75th	q25th
1998:2001	$8,656$	$30,470$	$24,894$	$-5,575$
2002:2005	$10,856$	$42,826$	$28,662$	$-14,164$
2006:2009	$3,737$	$48,689$	$26,428$	$-22,261$
2010:2012	$4,516$	$33,637$	$21,016$	$-12,622$

	Metropolitan Area of San Diego			
	Mean	q75th-q25th	q75th	q25th
1998:2001	$20,750$	$37,925$	$33,856$	$-4,069$
2002:2005	$26,608$	$71,930$	$52,806$	$-19,124$
2006:2009	$11,080$	$69,364$	$50,640$	$-18,724$
2010:2012	$5,873$	$45,267$	$31,421$	$-13,846$

	Metropolitan Area of San Francisco			
	Mean	q75th-q25th	q75th	q25th
1998:2001	$24,818$	$73,696$	$50,382$	$-23,315$
2002:2005	$18,674$	$109,246$	$77,906$	$-31,340$
2006:2009	$-6,209$	$119,949$	$60,508$	$-59,441$
2010:2012	$3,904$	$74,700$	$44,920$	$-29,779$

Table 17: Distribution of value added for house re-sales executed by active flippers. Abnormal performance is measured with respect to re-sales undertaken by non-active flippers. All values are in terms of December 2013 dollars.

	Metropolitan Area of Los Angeles			
	Mean	q75th-q25th	q75th	q25th
1998:2001	19,906	31,406	31,479	73
2002:2005	12,898	73,492	47,396	−26,096
2006:2009	50,511	82,223	86,162	3,939
2010:2012	4,251	53,194	33,460	−19,733
	Metropolitan Area of Sacramento			
	Mean	q75th-q25th	q75th	q25th
1998:2001	16,288	29,279	30,434	1,155
2002:2005	13,770	42,357	30,368	−11,989
2006:2009	16,896	48,359	37,718	−10,642
2010:2012	7,552	33,475	23,604	−9,871
	Metropolitan Area of San Diego			
	Mean	q75th-q25th	q75th	q25th
1998:2001	25,257	38,619	38,557	−62
2002:2005	42,583	69,377	66,936	−2,441
2006:2009	52,516	60,797	82,335	21,538
2010:2012	22,430	43,101	45,245	2,144
	Metropolitan Area of San Francisco			
	Mean	q75th-q25th	q75th	q25th
1998:2001	23,965	73,810	49,685	−24,125
2002:2005	19,488	109,221	78,523	−30,698
2006:2009	42,536	115,771	92,355	−23,415
2010:2012	3,436	74,558	44,439	−30,119

Table 18: Distribution of value added for house re-sales executed by active flippers. Abnormal performance is measured with respect to re-sales undertaken by non-active flippers. All values are in terms of December 2013 dollars.

The dispersion of outcomes for individual transaction is very large, even when measured in terms of value added. The bottom quartile of trades earns negative value added, even when we measure abnormal performance with respect to non flippers. Overall, our results show that the dispersion of transaction-level performance is large.

However, we cannot yet say whether the dispersion that we see in the data is generated by transaction-level risk, or by heterogeneity at the level of different house flippers. In other words, we have not yet investigated whether there are good and bad flippers, persistently over or under performing the average performance of their peers.

Korteweg and Sorensen (2016) address a similar problem, when studying the persistence of performance for private equity firms that manage multiple funds over time. Their approach is to decompose fund-level abnormal returns into a component explained by firm random effects and an unexplained residual component. This decomposition is basically an analysis of variance (ANOVA) of abnormal returns. We use the same principle in our data, and run an ANOVA of $\widehat{LME}_a$ and $\widehat{VA}_a$ (respectively, gross capital gains and value added in excess of non-active flippers). Table 19 summarizes results from the decomposition for a pooled sample of transactions across all metropolitan areas. Panel (a) reports results for $\widehat{LME}_{j,a}$, while panel (b) reports results for $\widehat{VA}_{j,a}$. In our calculations we again select all re-sales by active traders who performed at least two transactions in each one of the four vintage windows: from 1998 to 2001, from 2002 to 2005, from 2006 to 2009 and from 2010 to 2012.

The coefficient $\sigma(pool)$ is the standard deviation of the abnormal performance measure (either $\widehat{LME}_a$ or $\widehat{VA}_a$); $\sigma(LSDV)$ captures the standard deviation determined by variation within owners, while $\sigma(RE)$ is the standard deviation determined by variation across owners, or, in other words, determined by owner random effects; $\sigma(RE)^2/\sigma(pool)^2$ is the fraction of total variance that is explained by differences in performance across owners[13]. We can see that variation across owners explains a relevant fraction of the dispersion in outcomes across trades. As a fraction of variance, owners random effects explain between 30% and 47% of the dispersion in the data when performance is measured in terms of abnormal returns. When performance is measured in terms of value added, random effects capture between slightly above 50% and slightly below 20% of variation. While these numbers are relatively high, there is still plenty of variation in the data that is left unexplained. If we compare the values of $\sigma(LSDV)$ with the mean performance from tables 15 and 17, we still conclude that dispersion in performance is very large when compared to average performance.

	Panel (a): Local Market Equivalents			
	1998:2001	2002:2005	2006:2009	2010:2012
$\sigma(pool)$	25.42 %	20.20 %	30.22 %	22.48 %
$\sigma(LSDV)$	18.47 %	16.21 %	23.08 %	18.70 %
$\sigma(RE)$	6.96 %	3.99 %	7.14 %	3.79 %
$\sigma(RE)^2/\sigma(pool)^2$	47.24 %	35.62 %	41.69 %	30.85 %
	Panel (b): Value Added			
	1998:2001	2002:2005	2006:2009	2010:2012
$\sigma(pool)$	\$ 59,587	\$ 94,870	\$ 96,885	\$ 60,372
$\sigma(LSDV)$	\$ 41,548	\$ 80,546	\$ 87.426	\$ 52,926
$\sigma(RE)$	\$ 18,039	\$ 14,324	\$ 9,459	\$ 7,446
$\sigma(RE)^2/\sigma(pool)^2$	51.38 %	27.92 %	18.57 %	23.15 %
N Obs	1,196	1,771	1,784	8,139
N Traders	370	624	477	1,646
Mean Obs per Trader	3.2	2.8	3.7	4.9

Table 19: ANOVA for measure of abnormal performance of house re-sale; data from all metropolitan areas are split in subperiods based on vintages. Sample of trades performed by active traders for which we observe at least two transactions in any vintage subperiod.

The decision to pool performance measures across metropolitan areas does not appear to distort the main conclusions in our analysis. In table 20 we implement ANOVA on abnormal performance for

[13] In formulas, for example when considering $\widehat{LME}$, we calculate the pooled variance as $\sigma(pool)^2 = \frac{\sum_{j=1}^{J}\sum_{t=1}^{T_j}\left(\widehat{LME}_{j,t}-\overline{\widehat{LME}}\right)^2}{\left(\sum_{j=1}^{J}T_j\right)-K}$ where $\overline{\widehat{LME}}$ is the sample mean of the residual LMEs for the selected trades. J is the number of traders selected for the exercise, while T_j is the number of trades per trader j. Then, the LSDV variance is $\sigma(LSDV)^2 = \frac{\sum_{j=1}^{J}\sum_{t=1}^{T_j}\left(\widehat{LME}_{j,t}-\overline{\widehat{LME}_j}\right)^2}{\left(\sum_{j=1}^{J}T_j\right)-J-K}$, where $\overline{\widehat{LME}_j}$ is the average residual LME for trader j. The contribution of random trader effects to the ANOVA is then obtained as $\sigma(RE)^2 = \sigma(pool)^2 - \sigma(LSDV)^2$.

transactions taking place in the metropolitan area of Los Angeles. We choose this metropolitan area since it offers the largest cross-section of active traders. Results are qualitatively similar to the ones in table 19.

	Panel (a): Local Market Equivalents			
	1998:2001	2002:2005	2006:2009	2010:2012
$\sigma(pool)$	24.43 %	20.22 %	30.04 %	21.85 %
$\sigma(LSDV)$	16.97 %	15.42 %	22.33 %	17.68 %
$\sigma(RE)$	7.46 %	4.80 %	7.71 %	4.16 %
$\sigma(RE)^2/\sigma(pool)^2$	51.73 %	41.83 %	44.75 %	34.49 %
	Panel (b): Value Added			
	1998:2001	2002:2005	2006:2009	2010:2012
$\sigma(pool)$	$ 47,833	$ 91,793	$ 100,282	$ 61,486
$\sigma(LSDV)$	$ 25,653	$ 71,014	$ 91,726	$ 52,085
$\sigma(RE)$	$ 22,180	$ 20,779	$ 8,557	$ 9,401
$\sigma(RE)^2/\sigma(pool)^2$	71.24 %	40.15 %	16.34 %	28.24 %
N Obs	1,012	1,184	1,062	4,438
N Traders	298	434	291	901
Mean Obs per Trader	3.4	2.7	3.6	4.9

Table 20: ANOVA for measure of abnormal performance of house re-sale; data from the all metropolitan area of Los Angeles are split in subperiods based on vintages. Sample of trades performed by active traders for which we observe at least two transactions in any vintage subperiod.

7 Concluding Remarks

Illiquidity and segmentation are important features of housing markets, as shown by Piazzesi and Schneider (2009), Piazzesi et al. (2014) and several other studies reviewed by Han and Strange (2014). Asset dealers acting as middlemen play an important role in many other decentralized markets for real assets (see for example the work on commercial aircrafts by Gavazza (2011a) and Gavazza (2011b)). The presence of dealers in these markets has multiple general equilibrium implications. Dealers can lower trading delays by matching low and high valuation agents (see Gavazza (2015)) or by removing asymmetric information (see Biglaiser (1993), Boehmer et al. (2015)). It is therefore important to understand the role of asset dealers in residential real estate, and the extent to which they can extract economic rents and improve market liquidity.

In this paper, we use data from the main metropolitan areas of California (Los Angeles, Sacramento, San Diego and San Francisco) to study the behavior and performance of middlemen in housing markets. Our dataset is novel, since it provides information on both sale prices and remodeling expenses at the level of individual real estate properties. We therefore can keep track of the full sequence of investment cashflows for a house in between re-sales.

We focus our analysis on owners actively transacting houses over horizons shorter than one year, and therefore behaving as middlemen in the housing market. Using methodology from the private equity literature, we measure the abnormal returns earned by the transactions carried out by the middlemen. We show that, on average, these traders earn substantial abnormal capital gains.

However, the dispersion in the performance of individual transactions is large. Transaction-level risk affects middlemen, since the scale of their operations is small. Each middlemen performs only a small number of house re-sales at the same time. A possible explanation for transaction-level risk is that it could be mostly determined by heterogeneity in the skills of individual middlemen. However, we use variance decomposition analysis to show that variation across middlemen can explain only a limited fraction of the dispersion in the performance of individual trades.

We believe that our findings on middlemen performance can help explain why asset dealers play a relatively smaller role in housing rather than in other decentralized markets. Large scale of operations is needed in order to diversify away transaction-level risk. However, most players end up working at a small scale. Thus, to stay in the business, middlemen need to carry a large amount of transaction-level risk in their portfolio. On the other hand, the finding that house flippers on average can extract rents, suggests that a large intermediary with extensive resources could deliver value by stepping into local housing markets as an asset dealer.

In future work, we plan to develop a quantitative model to assess how frictions constraining the size of asset dealers could have led to the equilibrium that we observe in the data. Two frictions in particular may play a major role: access to capital and limited attention. The latter one is likely important in an environment with imperfect information. Limited attention could prevent flippers from monitoring a large enough area of the market or to inspect a large enough number of houses.

References

Bayer, Patrick, Christopher Geissler, and James W. Roberts, "Speculators and Middlemen: The Role of Flippers in the Housing Market," *Working Paper*, 2012.

Berk, Jonathan and Jules van Binsbergen, "Measuring Skill in the Mutual Fund Industry," *Journal of Financial Economics*, 2014.

_ and Richard Green, "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy*, 2004.

Biglaiser, Gary, "Middlemen as Experts," *RAND Journal of Economics*, 1993.

Boehmer, Ekkehart, Kingsley Y. L. Fong, and Wu Juan, "International Evidence on Algorithmic Trading," *Working Paper*, 2015.

Fama, Eugene and Kenneth French, "Luck Versus Skill in the Cross Section of Mutual Fund Returns," *Journal of Finance*, 2010.

Gavazza, Alessandro, "Leasing and Secondary Markets: Theory and Evidence from Commercial Aircraft," *Journal of Political Economy*, 2011.

_ , "The Role of Trading Frictions in Real Asset Markets," *American Economic Review*, 2011.

_ , "An Empirical Equilibrium Model of Decentralized Asset Markets," *Working Paper*, 2015.

Han, Lu and William C. Strange, "The Microstructure of Housing Markets: Search, Bargaining and Brokerage," in Duranton, Henderson, and Strange, eds., *Handbook of Regional and Urban Economics, Volume 5*, 2014.

Haughwout, Andrew, Donghoon Lee, Joseph S. Tracy, and Wilbert Van der Klaauw, "Real Estate Investors, the Leverage Cycle, and the Housing Market Crisis," *Working Paper*, 2011.

Kaplan, Steven and Antoinette Schoar, "Private Equity Performance: Returns, Persistence and Capital Flows," *Journal of Finance*, 2005.

Korteweg, Arthur and Morten Sorensen, "Skill and Luck in Private Equity Performance," *Journal of Financial Economics*, 2016.

Piazzesi, Monika and Martin Schneider, "Momentum Traders in the Housing Market: Survey Evidence and a Search Model," *American Economic Review*, 2009.

_ and _ , "Housing and Macroeconomics," in John Taylor and Harald Uhlig, eds., *Handbook of Macroeconomics*, Elsevier, 2016.

_ , _ , and Johannes Stroebel, "Segmented Housing Search," *Working Paper*, 2014.

_ , _ , and Tim Landvoigt, "The Housing Market(s) of San Diego," *American Economic Review*, 2015.

Sagi, Jacob S., “Asset-Level Risk and Return in Real Estate Investments,” *Working Paper*, 2015.

A Data Methodology

A.1 CoreLogic Deed Data

In this subsection, we detail the steps taken to organize the raw transaction data obtained from CoreLogic deed files from California, which are used to analyze the transactions returns in our results.

The first important thing to consider is the fact that multi-APN (Assessor Parcel Number) group sales hold an observation *per APN*. Since CoreLogic did not provide a multi-APN identifier, we classify as group sales whenever two or more data entries display the same record date, sale date, owner and seller names, document type, sale amount, mortgage amount and transaction type (whether it is resale or construction, excluding refinances) and *more than one APN*[14]. Moreover, we also classify as group sales if the data entries display a small time window of at least 5 days (and all the other information identical). Any identified group sale is removed from the exercise.

Secondly, we treat (and remove) duplicated observations. We developed a variable that flags potential duplicated observations in sample. Below we describe its groups:

- "Dup_flag" = -3 for true duplicates, which have all fields except for the same
- "Dup_flag" = -2 for almost true duplicates, which have same property id, date, legal description, buyer, seller, price, tax, document type, deed type, transfer type
- "Dup_flag" = -1 for properties with no APN
- "Dup_flag" = 0 for non-duplicates
- "Dup_flag" = 1: A transaction (possibly with a loan) and then one or multiple refinances
- "Dup_flag" = 2: A multiple refinance
- "Dup_flag" = 3: Simultaneous transfers, all transactions (before multiple buyers and sellers)
- "Dup_flag" = 4: Simultaneous transfers, not all transactions (before multiple buyers and sellers)
- "Dup_flag" = 5: Multiple buyers, same seller (some can be refinances, subdivision or construction))
- "Dup_flag" = 6: Multiple sellers, same buyer (some can be refinances, subdivision or construction))
- "Dup_flag" = 7: Chain of sales, transfers, and refinances, all sales same price
- "Dup_flag" = 8: Chain of sales, transfers, and refinances, 1 true sale
- "Dup_flag" = 9: Set of Sales With Same Buyer and Seller
- "Dup_flag" = 10: Pure Middle Man (Pair)

[14] If the APN remains the same, we classify the observations as duplicated, and therefore we keep only one observation as a transaction record.

- "Dup_flag" = 11: Other (includes chain of sales, transfers, and refinances with multiple true sales at different prices)

Whenever the flag indicates true or almost true duplicates, we keep only one observation per record. For all the other cases, we exclude potential duplicates (and cases when the APN is not available). We also remove refinances and equity transactions, and whether the transaction is part of a foreclosure process (notice of default for example).

For transfers (including non-arms length transactions, nominal transfers), we keep such observations in order to properly identify whether a return was generated over arms-length transactions or not. After comparing repeated sales, we remove any returns involving transfers.

Lastly, we opted to remove condos as they are not properly identified within a specific APN in a considerable part of the sample. Our sample concentrates specifically on single-family residences.

A.2 Merging Deed and Tax Assessment Data

We capture house specific information from tax assessment files provided by CoreLogic. We match tax files between housing units using the unformatted version of the APN (as provided by the county/city clerk) and the FIPS county/state code. Tax assessment files are available as of 2014 for the majority of the data.

A.3 Merging Corelogic and Buildzoom

At this stage, we match the data between CoreLogic and Buildzoom by comparing address location under the same city. Due to potential misreporting/misspelling of city names, we first geocode the lat/lon coordinates using a GIS software (ArcGIS, provided by Stanford University) and collect the city names (and 5-digit zipcodes) as reported in the Census of 2010[15]. When comparing properties that were remodeled against those that were not, we keep only cities in which we observe at least one remodeling observation in sample (other cities are excluded for potential coverage issues with Buildzoom).

In order to match Buildzoom information with CoreLogic, we compare addresses (street number and name only). We consider a match if:

- Street numbers are identical.
- Street names are no further than a normalized Levenshtein distance of 0.2[16].

Secondly, we repeat the same steps before for unmatched observations using the city names provided by Buildzoom. We assume that all matched information from the first stage should not be readjusted (since addresses are matching), therefore focusing only on the remaining unmatched fraction.

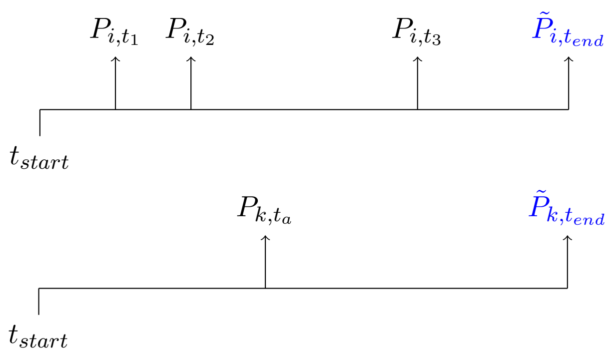
[15] Shapefile source: nhgis.

[16] The normalized Levenshtein distance is equal to the number of character changes needed to adjust in order to obtain a match divided by the joint length of the strings in analysis.

B Pricing Housing Units at the End of Sample

Consider properties i and k in our dataset. The diagram below shows the sequences of trades involving the two properties from the first available date, t_{start} (which is January 1998) onwards. For each transaction time we report the amount P at which the property was bought. We can see that property i was involved in three transactions at times t_1, t_2 and t_3, while property k is involved in a single transaction at time t_a. After respectively the transactions taking place at t_3 and t_a, the properties are held in the portfolio of their last buyer.

A first approach to the analysis of real estate returns would be to just focus on the transactions that have been completed. However, this might introduce bias in our results, since the decision to hold onto some properties might be endogenous to an assessment of market conditions on the side of real estate traders. It is therefore worth to include these transactions and try to come up with an estimate of their ongoing performance up to the end of sample date, t_{end} (which is December 2013). Note that in the diagram below we report for both properties a value $\tilde{P}$ on the timeline at time t_{end}. This value is needed in order to determine the performance of the last trades.



We therefore need to take a stance on properties valuation as of December 2013. We do so by running an hedonic regression on the log prices of houses transacted over the period between January and December 2013.

$$\log\left(P_{i,t}\right) = a_z + a_q + \mathcal{B}^{house} X_{i,t}^{house} + \mathcal{B}^{trade} X_{i,t}^{trade} + \mathcal{B}^{owner} X_{i,t}^{owner} + \epsilon_{i,t}$$

The dependent variable is the log price of property i at time t. The coefficients a_z and a_q are respectively ZIP code and quarterly fixed effects. The variable $X_{i,t}^{house}$ is a column vector and collects multiple house characteristics, such as number of bedrooms, number of bathrooms, square foot size of the house and other features; $\mathcal{B}^{house}$ is a row vector of coefficients for these characteristics. In the same fashion, $X_{i,t}^{trade}$ collects characteristics of the trade, for example whether it is a distress sale. Finally, $X_{i,t}^{owner}$ contains characteristics of the seller, such as whether the seller is a corporation or an LLC. Coefficients from the estimation are reported in the following table. We use the results from the hedonic regressions to price properties in the portfolio of owners, by matching seller characteristics with the characteristics of the owner.

	(Log) House Price
Age	-0.002∗∗∗
	(-9.03)
Acres	0.000∗∗∗
	(3.51)
N. Bedrooms	0.035∗∗∗
	(8.58)
Total Building Sq. Ft.	0.000∗∗∗
	(33.72)
N. Bathrooms	-0.001
	(-0.18)
N. Rooms	0.001
	(0.88)
Has Central A.C.	0.016∗∗
	(2.20)
Has Central Heating	0.010
	(1.57)
Depth Footage	0.000∗∗∗
	(3.66)
Has Fireplace	0.052∗∗∗
	(11.55)
Has Garage	0.059∗∗∗
	(3.45)
Has Parking Space	-0.029∗∗∗
	(-4.14)
Has Pool	0.066∗∗∗
	(14.89)
Low Accessed Quality	-0.058
	(-1.07)
Origination Loan Value	0.000∗∗∗
	(8.41)
Origination LTV ratio	-0.015
	(-0.94)
Construction (at purchase)	0.048∗∗∗
	(6.75)
Construction (at resale)	0.028∗∗∗
	(3.73)
REO (at sale)	-0.133∗∗∗
	(-30.24)
Short sale (at sale)	-0.192∗∗∗
	(-43.27)
Lender is Seller	-0.083∗∗∗
	(-6.76)
Buyer is Corporate	0.037∗∗∗
	(7.86)
Buyer is Investor	0.019∗∗∗
	(2.80)
Investor x INC	-0.177∗∗∗
	(-21.57)
Investor x LLC	-0.123∗∗∗
	(-15.93)
Investor x C-corp	-0.148∗∗∗
	(-10.49)
Passive Trader	0.040∗∗∗
	(15.74)
Zipcode FE	Yes
Year-Qt FE	Yes
Controls	Yes
R-sq	87.50 %
N	209,561

Table 21: Coefficients from hedonic regression. The regression is estimated over all house sales in 2013.