Idiosyncratic Risk in Housing Markets[*]

Marco Giacoletti[†]

Abstract

Most US households invest a substantial fraction of their wealth in a single asset: their family home. I study the empirical properties of idiosyncratic housing risk and its welfare cost from the perspective of homeowners. Using micro-data from the metropolitan areas of Los Angeles, San Diego, and San Francisco, I show that the fraction of house price standard deviation determined by idiosyncratic risk is decreasing as the holding period increases. Idiosyncratic risk makes up to 60% of capital gains standard deviation of home resales after one year, while it accounts for less than 20% after five years. The welfare cost of house-specific risk is assessed using a calibrated dynamic portfolio model. The cost is measured as the premium that homeowners would pay to insure against idiosyncratic risk. I find that annual premiums can be larger than one month of rent. The magnitude of the premiums critically depends on the likelihood of moving.

1   Introduction

Housing is the largest asset in the portfolio of most US households. Houses are by their nature a lumpy investment and most households own a single home. Thus, idiosyncratic house risk cannot be diversified away and has major implications for households and their economic welfare. The behavior of house-specific risk and its welfare cost from the perspective of homeowners are the focus of this project.

The first part of the paper is devoted to estimating idiosyncratic risk and studying its properties. I use novel data on repeated house sales from the metropolitan areas of Los Angeles, San Diego, and San Francisco. I compute the idiosyncratic standard deviation of capital gains for homes sold in each year from 2001 to 2012 and within each metropolitan area. Consistently with previous research, I find that idiosyncratic standard deviation is large relative to the annual volatility of metropolitan-area price indexes. Moreover, I show that idiosyncratic standard deviation has increased in all metropolitan areas during and after the housing crisis of 2007-2009.

The key empirical finding is that the magnitude of idiosyncratic risk does not depend on the length of the investment horizon. In other words, house-specific standard deviation does not scale with the (square root of the) holding period. Thus, the behavior of idiosyncratic risk in housing stands in sharp contrast to evidence from public equity markets, where idiosyncratic stock volatility scales with the square root of time. Unlike idiosyncratic risk, the standard deviation of metropolitan-area house price fluctuations is increasing in the investment horizon. Thus, the relative contribution of idiosyncratic risk to house price standard deviation changes with the holding period. Idiosyncratic risk can make up to 60% of the standard deviation of capital gains when a home is resold after one year. However, after five years, the idiosyncratic component is at most 20% of total standard deviation.

The results on the term structure of idiosyncratic risk also shed light on the nature of house-specific shocks. If house-specific shocks were generated by a sum of *i.i.d.* components accumulating over time, their magnitude would scale with the investment horizon. In the data, house-specific risk is more likely to be determined by a single shock, whose size does not depend on the holding period. In fact, one can think that the shock is realized at the time of sale. There is a broad body of literature that documents how housing markets are illiquid and segmented (see the work on search and segmentation of Piazzesi and Schneider (2009b), Piazzesi et al. (2014) and several other studies reviewed by Han and Strange (2014)). The primary driver of house-specific risk might not be a change in house quality or physical characteristics, but rather price uncertainty due to market inefficiency.

In the second part of the paper, I use a quantitative portfolio model to assess the impact of idiosyncratic risk on household welfare. To my knowledge, this paper is the first to quantify the

cost of house-specific risk. I measure the welfare cost as the annual premium that a homeowner would be willing to pay to insure against house-specific shocks. I find that premiums are sizable and can exceed the value of one month of rent.

In the model, a homeowner may sell her house over a finite planning horizon. A home sale can be determined either by an optimal decision or by an exogenous mobility event that forces the household to move. In line with my findings from the empirical section, I model idiosyncratic risk as a one time shock that affects the house price at the time of sale. This implies that the homeowner has no information on house-specific shocks until the house is sold. The setup is a reasonable approximation to the data. Houses are usually valued only when a sale listing is prepared. Furthermore, house prices are only observable when home sales take place. Finally, the idiosyncratic shock may be realized at the time of sale.

The dynamics of the "systemic" component of house prices are calibrated to match the ones of metropolitan-area house price indexes. Real housing services are kept constant over the entire holding period of the house. In the model, after a house sale takes place, the household buys a new home in the same metropolitan area. The capital gains and the ratio of price to housing services for the new house are perfectly correlated with the ones of the original home, with the exception of the idiosyncratic shock. This is a simplification[1], broadly consistent with evidence from the data. According to the US Census, most household mobility takes place within the same metropolitan area or the same state. Metropolitan-area house price indexes are highly correlated within California.

In the context of the model, I study how insurance premiums change when I assign to homeowners different probabilities of facing exogenous mobility shocks and different amounts of initial non-housing (financial) wealth. *Ceteris paribus*, higher mobility leads to higher insurance premiums since idiosyncratic risk has a greater impact on house capital gains for shorter holding periods. Households who have greater financial wealth can use it as a buffer to absorb idiosyncratic house price shocks. This is particularly important when the homeowner is forced to sell due to exogenous mobility. The model is calibrated to take into account life-cycle effects on income for households of different ages.

Insurance premiums are the largest for young households that have small non-housing wealth and a high likelihood of being hit by mobility shocks. These households are willing to pay annual premiums that exceed one month of rent. The impact of exogenous mobility shocks is large. When household age is in between 25 and 29, increasing the probability of exogenous mobility shocks from 0% to 30% on an annual basis more than doubles insurance premiums for homeowners with

---

[1]The perfect correlation (excluding the idiosyncratic shock) between price fluctuations of the first and the new house maximizes the hedging benefits of home ownership. In fact, the house (again, excluding the idiosyncratic shock) perfectly hedges the household against future house prices or rent expenses. For more details on the hedging effects of home ownership, see Sinai and Souleles (2005), Lu (2010) and Lu (2013).

non-housing wealth below $50,000.

My work brings novel contributions to several branches of the financial economics literature on housing. The data and methodology that I use to measure idiosyncratic housing risk are novel. My empirical analysis is based on a unique dataset of home resales built by Giacoletti and Westrupp (2016). The data are constructed by merging sales records from Corelogic deed files with data on remodeling permits collected by Buildzoom[2]. By including remodeling expenses, I follow the entire stream of investment cashflows that takes place in between house resales[3].

Measuring idiosyncratic house price risk with sales micro-data is challenging for several reasons. The performance of individual transactions is observed only at the time of sale. Furthermore, due to the inclusion of remodeling expenses, each deed can involve multiple cashflows. Similar issues emerge in assessing the performance of private equity funds. Kaplan and Schoar (2005) introduce the *public market equivalent* (PME) as a measure of abnormal performance of a fund on an investment in the S&P 500 with the same cash inflows. In a similar manner, Giacoletti and Westrupp (2016) develop the *local market equivalent* (LME) to measure abnormal returns on ZIP-code level house price indexes.

In this paper, I compute local market equivalents on metropolitan-area indexes. I measure the local market equivalent as the difference between the house sale price and the terminal value of an investment with the same cash inflows, capitalized at the rate of return on the local metropolitan-area house price index. The local market equivalent would already be a satisfactory measure of abnormal capital gains only if the loading of metropolitan-area market fluctuations on each house was equal to one or, in other words, only if each home had a beta of one with respect to index returns. However, house capital gains are determined by local factors, as well as by physical characteristics of the house. Moreover, Piazzesi et al. (2015) show that, during the years of the housing boom, lower quality (cheaper) houses systematically appreciated more than higher quality homes.

These effects are taken into account by my empirical strategy. I group resales by year of sale and run regressions of local market equivalents on several control variables. These account for differences in location, as well as house and sale (houses sold through short sales experience substantial price discounts) characteristics. As already pointed out above, I verify that residuals from these regressions are not locally clustered. This ensures that there are no omitted local factors confounding my estimates of house-specific risk. Eventually, idiosyncratic risk is computed as the

---

[2]Buildzoom is an intermediary in the remodeling market for residential and commercial real estate. I thank Issi Romen and Buildzoom for making the data available.

[3]Investment in remodeling has to be included in the calculations of individual home returns, since it has an impact on resale prices. The presence of a non-temporal components in house capital gains, potentially driven by remodeling activity, is first acknowledged by Goetzmann and Spiegel (1995). The effects of remodeling cashflows on house capital gains are then documented by Knight and Sirmans (1996) and Wilhelmsson (2007) with sales data, as well as by Harding et al. (2006) with survey evidence from the American Housing Survey (AHS).

dispersion of regression residuals within a metropolitan area.

My findings on the relationship between idiosyncratic risk and holding period, are related to previous research on housing and other illiquid asset classes. In fact, there is previous empirical work documenting the dispersion in prices and capital gains for individual houses within a metropolitan area. Seminal work by Case and Shiller (1988) highlights that prices of individual homes are twice more volatile than city-level house price indexes. Similar findings are later reported by Flavin and Yamashita (2002), Piazzesi et al. (2007) and Piazzesi et al. (2015). This last paper also shows evidence hinting that idiosyncratic risk has increased during the housing crisis.

With the exception of Case and Shiller (1988), all the papers just mentioned assume that idiosyncratic risk scales with time. While this assumption is standard in the most recent literature, early papers in housing had a more articulated view of the behavior of idiosyncratic housing risk. Case and Shiller (1987) and Goetzmann (1993) document that the standard deviation of individual house capital gains does not converge to zero as the holding period shrinks. This result suggests that idiosyncratic housing risk can not be modeled only has a sum of *i.i.d.* shocks, but rather a transitory component is also needed.

However, the papers mentioned above do not investigate in detail the properties of idiosyncratic risk across different investment horizons[4]. In my study, I not only show that there is a transitory component in idiosyncratic risk, but I argue that this component is the dominant driver of house-specific risk. I show that there is no evidence of a persistent component (scaling with time) even when considering longer holding periods.

Empirical results consistent with the ones emerging from my work can be found in another older study, this one based on Swedish data, by Englund et al. (1998). The paper uses micro-data for house sales taking place in Sweden over the years from 1980 to 1993. After controlling for house characteristics in hedonic regressions, the authors estimate an autoregressive process for idiosyncratic house shocks at quarterly frequency. If the autoregressive coefficient were equal to one, idiosyncratic risk would be scaling with the holding period. The authors find an autoregressive coefficient of -0.6, which is consistent with very low autocorrelation over horizons of multiple years.

The findings in my paper, based on data from California and years from 1990 to 2012, are even more clearly in favor of the lack of any holding period dependence in the structure of house-specific shocks. In my empirical work, I go to greater length than the previous literature to identify the idiosyncratic component of house capital gains. My aim is to provide reliable estimates of house-specific shocks. Thus, I control for variation that can be explained by metropolitan-area risk,

---

[4]Moreover, their empirical evidence is largely based on data from the 1970s and 1980s. Information efficiency and conditions in housing markets are likely to have substantially changed over time. Case and Shiller (1987) work with data from US cities (Atlanta, Chicago, Dallas and San Francisco), and Goetzmann (1993) bases his analysis on the results from Case and Shiller (1987).

local (ZIP-code level) fluctuations and house characteristics. To ensure that the effects of local market factors are fully removed, I verify that the estimated house-specific shocks are not clustered. To this end, I show that the shocks have zero spatial auto-correlation. In other words, there is no correlation of estimated house-specific shocks among houses that are geographically close.

This topic of research is not restricted to housing. There is work on other illiquid asset classes that studies the behavior of idiosyncratic risk as a function of the holding period. Axelson et al. (2015), who study leveraged buyout funds in private equity, find that idiosyncratic risk does not converge to zero as the holding period shrinks. A similar finding is also reported by Peng (2015) and Sagi (2015) for commercial real estate. Sagi (2015) shows, using a quantitative model (based on the framework developed by Duffie et al. (2005)), that the transitory component of idiosyncratic risk can be generated by trading in illiquid markets, where sellers bear transaction cost, and asset valuations are persistent but heterogeneous. While I believe that this explanation is appealing also for housing, it is hard to identify the determinants of idiosyncratic risk in the data without information on the individual house listing processes. Thus, in the current paper I am agnostic about the sources of house-specific risk, and defer this topic to further research.

The model developed in this paper also delivers novel contributions. A few quantitative models in the housing literature have included idiosyncratic housing risk as an element in their structure. However, my paper is the first to focus its analysis on the cost of house-specific risk. Moreover, previous models have been calibrated so that idiosyncratic housing risk scales with the investment horizon. This is the case for Flavin and Yamashita (2002), for Favilukis et al. (2010) and for Piazzesi et al. (2015).

The quantitative portfolio model in this paper is inspired by the frameworks in Campbell and Cocco (2003), Cocco (2005b) and more recently Chen et al. (2013). The homeowner consumes a basket of numeraire good and housing services and earns an income stream subject to persistent idiosyncratic shocks. The menu of financial assets available to the household is limited to a risk-free bank account. I consider both a version of the model where the household owns the entire capital invested in her house and a version where the household is financing her house with a nominal fixed rate mortgage (FRM). The dynamics of prices, interest rates, rent rates and income shocks are exogenously determined. My model is separately solved for households of different ages. Optimal decisions are taken over a planning horizon of five years, at the end of which the household receives terminal wealth as consumption. In the calibration, I take into account life-cyle effects. Income level and expected income growth are estimated for households of different ages using data from the Current Population Survey for California. Financial wealth accumulation at different stages of the life-cycle is calibrated to match data from the Survey of Consumer Finances.

My work is also related to papers that study the relationship between house price fluctuations and consumption for young and old households (from the seminal contribution of Campbell and

Cocco (2007) to the more recent work of Gan (2010), Iacoviello (2011), Attanasio et al. (2011) and Badarinza et al. (2016)). Moreover, it is related to work that focuses on the relationship between house price fluctuations and household portfolio decisions (Yao and Zhang (2005), Cocco (2005a) and the empirical work in Chetty et al. (2016)). I show that house-specific risk is costly for young households. Since young households are the ones most likely to be entering the housing market for the first time, idiosyncratic house risk might critically affect their choice between owning and renting.

The rest of the paper proceeds as follows. A first part is devoted to the description of the micro-data, the estimation methodology and the empirical findings. This consists of sections from 2 to 7. Section 2 focuses on the repeated house sales micro-data. Section 3 discusses in detail the methodology used to measure house-specific shocks and provides estimates of idiosyncratic housing risk. Section 4 studies whether idiosyncratic risk changes across subsamples of house resales with different characteristics, and most importantly shows that idiosyncratic risk does not scale with investment horizon. Section 5 shows that the findings on the term structure of idiosyncratic risk are robust to different definitions of idiosyncratic shocks, and in different sub-samples of the data. Section 6 documents the relationship between idiosyncratic risk and investment horizon in equity markets. Section 7 investigates the relative importance of idiosyncratic and metropolitan-area risk for different holding periods. The second part of the paper is devoted to the calibrated model and starts with section 8. Section 8 introduces and describes the quantitative model, while section 9 discusses in detail its calibration. Section 10 reports the welfare costs and insurance premiums generated by the model. Finally, my concluding remarks are in section 11.

2   House Sales Data

The key empirical results in the paper are based on a unique database consisting of micro-data on both house repeated sales and remodeling permits that has been originally built by Giacoletti and Westrupp (2016). Information on house transactions is provided by Corelogic Solutions[5]. In the data, housing units are identified by their address and Assessor Parcel Number (APN). The latter is an identifier assigned to each plot of land by tax assessors for record-keeping purposes. The Corelogic "deeds" dataset provides information on transaction prices, sellers, buyers and the mortgage originated by the buyer at the time of the house sale. Corelogic also provides tax assessment data for the fiscal year 2013-2014, which contain geo-location information for each APN (in terms of latitude and longitude) as well as a snapshot of house characteristics (housing unit square feet size, number of bathrooms, number of bedrooms and other features). After merging

---

[5]Information on the data provider is available at http://www.corelogic.com/industry/real-estate-solutions.aspx

deeds and tax assessment information for each APN, I select house sales carried out as arm's length transactions and exclude nominal sales.

Information on remodeling activity in the state of California is provided by Buildzoom[6]. The company acts an intermediary matching owners and contractors in the commercial and residential real estate remodeling market. Buildzoom collects information on house remodeling and building permits from local Census authorities. The data report the kind of job that was performed (for example, kitchen renovation, plumbing or roof repair), the job cost and the fees paid by the contractor. Housing units are again identified by their address and APN. This makes it possible to merge the Buildzoom data with the Corelogic database. Details on the methodology and the issues faced in the merge can be found in appendix C. I organize the data so that I can keep track of repeated sales for the same housing unit. Thanks to the remodeling data, for each resale I can collect all investment flows, from the time when the house is bought to the time when the house is sold.

I focus my analysis on three main urban areas of California: the counties of Los Angeles and San Diego and the Metropolitan Statistical Area (MSA) of San Francisco (which includes Alameda county, Contra Costa county, Marin County, San Francisco County, San Mateo County) extended to include Santa Clara County. These are some of the most important housing markets in the United States, which are well covered both in Corelogic and the Buildzoom datasets. Furthermore, an argument can be made that California markets are ideal for studying house-specific risk. In fact, even though in my analysis I will be very careful in accounting for all factors that might explain systematic differences in returns between houses, it is still true that house structures are very heterogeneous. Some very specific features of house layouts might be hard to account for and might have an impact on house capital gains. However, in California the dominant determinant of house value is residential land value. This appears very neatly if we look at land value and house structure value indexes developed by Davis and Palumbo (2006) for all the main US cities. In figure B.24, using Davis and Palumbo (2006) data, I compute the average fraction of house value that is determined by land value within the metropolitan areas of Los Angeles, San Diego and San Francisco over the period from the first quarter of 1990 to the last quarter of 2012. Over the entire period, on average 70% of house value is determined by land value in Los Angeles and San Diego, and almost 80% in San Francisco. This is in contrast with the national average, where land value was 40% of house value before the housing crisis and only 30% thereafter.

In the final dataset, I exclude resales that are sold as REOs, both because it is hard to assess the total performance of these trades and because foreclosed homes frequently experience drops in their maintenance and conditions (see for example the evidence in Anenberg and Kung (2014)).

I also exclude all trades with holding period shorter than one year. Both Bayer et al. (2012) and

---

[6]Information on the company is available at https://www.buildzoom.com/about.

Giacoletti and Westrupp (2016) show that this segment of the market is dominated by middlemen and professional investors. Giacoletti and Westrupp (2016) show that transactions with holding period shorter than one year earn on average large abnormal returns with respect to local house price indexes. This is not the case for transactions with longer holding periods.

Transactions by agents that seem to be professional traders are quite rare. Giacoletti and Westrupp (2016) identify professional traders as either legal entities or individuals that have been recently involved in multiple house trades. Panel (a) of figure B.25 reports for each year from 2001 to 2012 the fraction of resales undertaken by individuals that bought more than one other house over the two years preceding the sale. The fraction is frequently smaller than 10 % and peaks in 2006 at 11.5%. Panel (b) shows the fraction of resales carried out by legal entities (including both corporations and trusts). This fraction is below 3 % for most years and reaches 5 % only at the end of the sample. From table 1, we can see that house resales represent approximately 20 % of all house sales available in my dataset each year. There is a concern that the distribution of resales might end up being clustered in specific geographies, delivering biased information on the metropolitan areas included in the study. Section D of the appendix shows in detail that this is not the case. Resales in my final dataset are widely spread across metropolitan areas and are present in both high and low sale density neighborhoods.

Table 1 reports sample statistics of the data for the years from 2001 to 2012, where I group house resales based on the year of sale. The column with header "All" shows the total number of house sales available for a specific year after discarding non arm's length transactions, nominal sales and incomplete records, as I explain in detail in appendix C. Column "All no REOs" shows the total number of sales excluding REOs.

The first column ("N Obs") shows the number of sales that I have successfully identified as resales and are included in the final database. For each resale, I was able to match the sale record with the record of the previous transaction with which the current seller became owner of the house. 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 from the nineties onwards. Moreover, coverage from the Buildzoom dataset is best only after 1998. Thus, I base resale transactions only on deed data from January 1990 onwards. Table 1 reports the fraction of resales that were "initiated" before 1998 and therefore potentially missing remodeling cashflows, as well as the fraction of resales that involved remodeling activity. While may be missing data on remodeling activity for the first few years, both the fraction of pre-1998 trades moves downwards and the fraction of remodeled homes goes upwards quickly.

	N Obs	Med HP	Med Price	Med LTV	Fr Pre-1998	Fr Rem	Fr Sh Sales	All no REOs	All
2001	21,961	3.42	382,372 $	90.00%	48.98%	4.97%	–	81,495	83,458
2002	29,602	3.67	434,272 $	90.00%	36.27%	6.18%	–	96,799	98,437
2003	32,351	3.83	486,590 $	90.00%	27.14%	7.76%	–	107,450	108,201
2004	33,655	3.83	577,998 $	90.00%	20.38%	8.48%	–	112,638	113,125
2005	33,510	3.58	660,098 $	90.00%	15.13%	9.27%	–	106,275	106,556
2006	27,088	3.33	680,431 $	90.00%	11.08%	10.22%	0.37%	84,164	84,620
2007	17,795	3.67	732,181 $	88.33%	9.15%	12.00%	3.25%	57,208	61,378
2008	13,095	3.92	597,194 $	90.00%	7.66%	12.33%	25.26%	43,141	74,025
2009	16,284	4.58	475,770 $	90.00%	6.69%	11.79%	35.76%	53,955	87,676
2010	18,245	5.33	468,098 $	89.96%	6.44%	13.01%	35.18%	62,292	83,871
2011	18,163	6.17	433,585 $	89.66%	5.96%	14.60%	37.32%	66,607	87,619
2012	22,477	6.75	423,495 $	87.50%	5.46%	15.08%	34.13%	76,481	91,046

Table 1: Sample statistics for house sale in years from 2001 to 2012. Columns 1 to 7 refer to resales, while the last two columns show the total number of sales without and with REOs.

3   Idiosyncratic Risk in House Sales

The aim of this section is to measure idiosyncratic risk in the metropolitan areas of Los Angeles, San Diego and San Francisco. I focus my analysis on houses sold in the period from January 2001 to December 2012. Due to data quality issues, I consider only resales of houses bought from January 1990 onwards.

As a first step, it is important to assess the magnitude and behavior of "aggregate" fluctuations that impact the entire metropolitan area. To this end, I use S&P Case-Shiller indexes[7] for the cities of Los Angeles, San Diego and San Francisco. Figure 1 reports the monthly time series of the indexes over the period between January 1990 and December 2013. I rescale the time series to reflect real returns[8]. Table 2 shows correlations among real monthly returns of the city-wide indexes, as well as monthly volatility and annual volatility (obtained by multiplying monthly volatility times $\sqrt{12}$) for each index. I consider both the entire sample (from January 1990 to December 2013) and a shorter sample starting on December 2001. Index returns are highly correlated across the three metropolitan areas. Estimated correlations are similar in the two samples and are always larger than 0.8 and larger than 0.9 in between Los Angeles and San Diego. Estimates of annual standard deviations are around 4% for the longer sample and around 5% for the most recent sample.

I measure idiosyncratic risk on top of index volatility using the micro-data on house resales described in section 2. These transactions involve multiple cash inflows, due to remodeling expenses. Kaplan and Schoar (2005) face similar issues when computing the abnormal returns of private equity funds in excess of a public equity benchmark. They introduce the Public Market Equivalent (PME), which measures the abnormal performance of a fund measured against a portfolio with the

---

[7]Available from the St. Louis Federal Reserve Bank database at https://research.stlouisfed.org/

[8]I deflate the price indexes using the consumer price index for all goods less food and energy provided by the US Bureau of Labor Statistics (more information can be found at http://stats.bls.gov:80/cpi/cpifaq.htm).
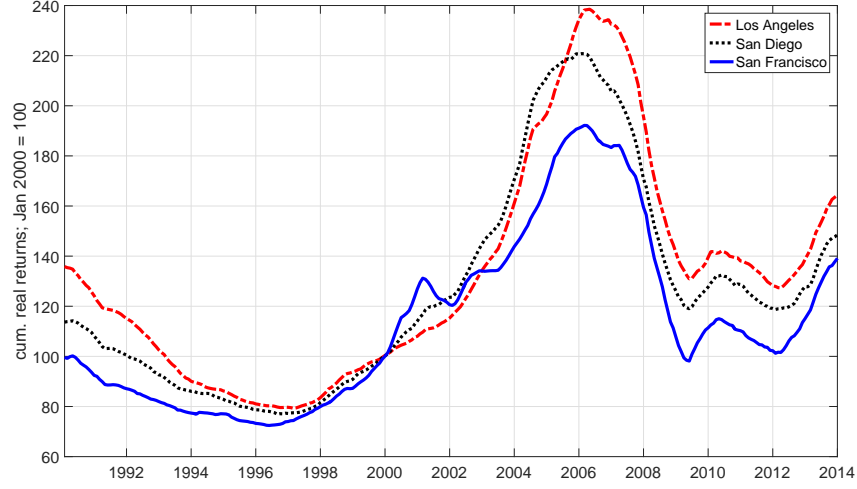
Figure 1: Real S&P Case-Shiller home price indexes for the cities of Los Angeles, San Diego and San Francisco. Monthly data from January 1990 to December 2013.

| | From 1990/01 to 2013/12 | | |
	Los Angeles	San Diego	San Francisco
Los Angeles	1		
San Diego	0.9087	1	
San Francisco	0.8068	0.8418	1
Std	1.17 %	1.13 %	1.28 %
Std Annualized	4.07 %	3.90 %	4.42 %

| | From 2001/01 to 2013/12 | | |
	Los Angeles	San Diego	San Francisco
Los Angeles	1		
San Diego	0.9222	1	
San Francisco	0.8452	0.8490	1
Std	1.41 %	1.37 %	1.51 %
Std Annualized	4.90 %	4.75 %	5.24 %

Table 2: Correlation matrix and standard deviation of real returns for S&P Case-Shiller indexes for the cities of Los Angeles, San Diego and San Francisco.

same inflows invested in the S&P 500. Giacoletti and Westrupp (2016) follow the same principle and compute the *Local Market Equivalent* (LME) to measure the abnormal return of a house trade with respect to a local house price index.

In this paper, I use a "forward" version of the local market equivalent, which measures the price premium (or discount) of a house with respect to an equivalent investment in the metropolitan-area

market index. The "forward" LME for house transaction i is computed as in equation 1:

$$LME_{fwd,i} = \frac{P_{i,T_i} - P_{i,T_i}^{loc}}{P_{i,T_i}^{loc}} \tag{1}$$

$$P_{i,T_i}^{loc} = \sum_{h=1}^{H_i} D_{i,\tau_{i,h}} R_{\tau_{i,h},T_i}^{Loc} + P_{i,t_i} R_{t_i,T_i}^{Loc}$$

With P_{i,t_i} being the nominal price at which the housing unit was bought, P_{i,T_i} the price at which the unit was sold, and $D_{i,\tau_{i,h}}$ are intermediate investments in home remodeling. P_{i,T_i}^{loc} is the nominal price of the "mimicking portfolio" invested in the metropolitan area index. The LMEs can then be consistently grouped by sale year. The cross-sectional dispersion of the LMEs for homes sold in a specific year is a first rough measure of idiosyncratic risk.

Note that the calculations in equation 1 do not explicitly account for rent income produced by the house and by the mimicking portfolio. The LME, as a measure of the total abnormal performance of the trade, is invariant to the inclusion of rents only when the house and the index investment produce the same rent payments. This is a problem, since it is likely that houses with different characteristics will earn different rent rates over time.

Moreover, as already pointed out by Kaplan and Schoar (2005), the PME (and by extension the forward LME in my calculations) is a measure of abnormal risk adjusted returns only if the loading or beta of returns for the individual funds (or, in my case, for the individual houses) with respect to the index is equal to one[9].

This assumption is likely to be violated at the level of individual house transactions. House capital gains are driven by local factors and house characteristics. Moreover, Piazzesi et al. (2015) and Piazzesi and Schneider (2016) show that the magnitude of house price fluctuations during the recent housing boom and bust were different for more and less expensive houses (which they use as a proxy for high and low quality houses). They also show that the relationship between price and house returns changed over time. Finally, it is reasonable that trade characteristics, such as the presence of a short sale, may determine discounts in home sale prices.

These problems are similar to the ones faced in the labor economics literature. In order to isolate idiosyncratic household income risk, researchers need to take into account household characteristics that may systematically lead to higher or lower income prospects (such as education, see Meghir and

---

[9]Korteweg and Nagel (2015) introduce the generalized public market equivalent (GPME), which weakens the assumption that beta is equal to one. However, GPME calculations are based on the additional assumption that beta is the same across all assets used in the calculation and that beta does change over time. The GPME basically treats the entire group of investments as a uniform asset class and identifies the beta using the overall timing of cashflows. If overall investment volume is pro-cyclical, the GPME will provide a beta estimate greater than one. If the investment volume is counter-cyclical, beta will be negative. While GPME is a valid methodology when studying private equity and venture capital as asset classes, it is not well suited for my settings, where I am focusing on individual house risk.

Pistaferri (2011) for a review of the literature). They do that by regressing relevant characteristics on labor income and taking the unexplained residuals as a proxy for idiosyncratic shocks. Along similar lines, I model the log LME so that its mean can depend on individual house and trade characteristics:

$$lme_{fwd,i} = \log\left(1 + LME_{fwd,i}\right)$$
$$lme_{fwd,i} = m(X_i) + \varepsilon_{fwd,i}$$

The conditional mean of the log LME $m(X_i)$ is estimated according regression equation 2:

$$m(X_i) = a_{sy\times zip}^{se(i)} + \mathcal{B}_{hchar}^{se(i)} X_{i,hchar} + \beta_{rem}^{se(i)} I_{rem,i} + \beta_p^{se(i)} \log(P_{i,t_i}) \tag{2}$$
$$+ \beta_{p,3y:6y}^{se(i)} \log(P_{i,t_i}) \cdot I_{i,3y:6y} + \beta_{p,>6y}^{se(i)} \log(P_{i,t_i}) \cdot I_{i,>6y} + \beta_{shrt}^{se(i)} I_{shrt,i}$$

The coefficient $a_{sy\times zip}^{se(i)}$ is a fixed effect specific to the year and ZIP-code in which the house was bought. The time and ZIP-code effects are meant to absorb differences in average house returns determined by local market fluctuations and across different holding periods. The vector $X_{i,hchar}$ contains house characteristics, in particular number of bedrooms, number of bathrooms and house size in squared feet. The dummies $I_{rem,i}$ and $I_{shrt,i}$ are equal to one respectively for houses that are remodeled and for houses that are sold through a short sale. The variable $\log(P_{i,t_i})$ is the log of the price at which the house was bought, which include as a proxy for house quality as done by Piazzesi et al. (2015).

The correlation between local market equivalent and log price might change based on the holding period of the house. I therefore introduce holding period dummies. In the final specification reported in 2, the dummy variables are $I_{i,3y:6y}$ and $I_{i,>6y}$. They are equal to one respectively when the holding period is in between three and six years and longer than six years. I have implemented different specifications where the dummy variables are defined over different holding periods. My results appear to be robust to the specification of these holding period dummies.

The holding period dummies are interacted with $\log(P_{i,t_i})$, so that the coefficient $\beta_p^{se(i)}$ captures the conditional correlation between log price and local market equivalent for houses with holding period up to three years, while $\beta_p^{se(i)} + \beta_{p,3y:6y}^{se(i)}$ captures the conditional correlation when the holding period is in between three and six years and $\beta_p^{se(i)} + \beta_{p,>6y}^{se(i)}$ captures the correlation when the holding period is longer than six years.

Equation 2 is separately estimated for houses sold in each year and in each metropolitan area. Coefficient estimates are reported in tables 3, 4 and 5. Consistently with the results in Piazzesi et al. (2015), I find that there is a negative relation between the price at which a house was bought and its abnormal performance with respect to the index. This result holds in all metropolitan

areas in the study, even though the coefficients on the log price and its interaction with the holding period dummies are sometimes not significant. Remodeling seems to deliver value for homeowners, since transaction involving remodeled houses generate on average positive abnormal returns. Short sales are on average settled at a significant (and economically large) price discount with respect to normal sales in all metropolitan areas.

	$IsRemodeled$	$\log(P)$	$\log(P) \cdot I_{3y:6y}$	$\log(P) \cdot I_{>6y}$	$IsShort$	R2	N Obs
2001	0.0472 [3.7234]	−0.2680 [−22.6980]	−0.0016 [−27.5479]	−0.0020 [−15.0131]	−	0.4550	9,299
2002	0.0392 [5.4087]	−0.2998 [−23.5004]	−0.0009 [−14.5863]	−0.0011 [−18.0981]	−	0.4452	12,972
2003	0.0423 [3.7764]	−0.3512 [−25.3366]	−0.0020 [−11.5815]	−0.0040 [−23.3737]	−	0.4899	14,384
2004	0.0211 [2.0490]	−0.4068 [−15.5199]	−0.0012 [−4.6444]	−0.0042 [−13.6177]	−	0.5191	14,433
2005	0.0182 [4.3162]	−0.4178 [−19.6118]	−0.0013 [−6.0120]	−0.0022 [−4.4931]	−	0.5774	15,113
2006	0.0234 [2.6189]	−0.4341 [−15.4817]	−0.0025 [−9.1802]	−0.0046 [−6.7701]	−0.2032 [−6.2229]	0.6316	12,806
2007	−0.0003 [−0.0282]	−0.3415 [−11.6626]	−0.0028 [−12.9902]	−0.0029 [−3.1729]	−0.1448 [−15.1386]	0.4946	7,453
2008	0.0282 [3.7220]	−0.1798 [−7.8933]	−0.0005 [−2.8500]	−0.0030 [−6.3015]	−0.0732 [−5.3735]	0.6001	4,969
2009	0.0136 [1.3283]	−0.1720 [−6.7243]	−0.0010 [−8.5547]	−0.0023 [−5.6933]	−0.0895 [−8.3025]	0.7223	6,667
2010	0.0008 [0.0787]	−0.2206 [−5.5694]	0.0007 [2.4495]	−0.0010 [−4.1933]	−0.1115 [−11.2413]	0.7536	7,331
2011	0.0381 [2.8883]	−0.2252 [−4.8045]	−0.0052 [−4.3450]	−0.0075 [−6.7566]	−0.1246 [−12.4409]	0.7255	7,131
2012	−0.0035 [−0.3179]	−0.2516 [−5.8151]	−0.0018 [−4.7769]	−0.0030 [−8.1964]	−0.1482 [−10.5370]	0.7247	8,453

Table 3: Coefficient estimates for year by year regressions based on equation 2 for the city of Los Angeles. T-stats are reported in brackets and are based on standard errors clustered by the year the transaction was initiated and by census tract.

I use the residuals from regression equation 2 as a proxy for idiosyncratic shocks to house prices. As a sanity check, I compute the spatial correlation of these residuals in each year and in each city. My aim is to make sure that variation in residuals is not driven by other important spatial-economic factors omitted from the conditioning information in equation 2. Spatial correlations are measured using Moran's I[10]. The Moran's I coefficient can assume values in between 1 and -1. Positive spatial correlation implies clustering of residuals. In my data, this would hint at the existence of local omitted factors. Results for each year and city are reported in table 6. We can see that Moran's I estimates are always in a range between -0.1% and 1.5% and therefore extremely close to zero.

---

[10]This measure of spatial correlation was first introduced by Moran (1950). It is computed as:

$$I = \frac{N}{\sum_i \sum_{j \neq i} w_{i,j}} \frac{\sum_i \sum_{j \neq i} w_{i,j} \left(X_j - \bar{X} \right) \left(X_i - \bar{X} \right)}{\sum_i \left(X_i - \bar{X} \right)^2}$$

Where N is the total number of observations, X_i is the value of the variable of interest for observation i, $\bar{X}$ is the sample mean of the variable we are interested in. The weights are $w_{i,j} = 1/d_{i,j}$, with $d_{i,j}$ the distance between i and j.

	$IsRemodeled$	$\log(P)$	$\log(P) \cdot I_{3y:6y}$	$\log(P) \cdot I_{>6y}$	$IsShort$	R2	N Obs
2001	0.0204 [4.3669]	−0.2491 [−16.0344]	−0.0010 [−8.7846]	0.0004 [1.4507]	−	0.3942	3,706
2002	0.0230 [3.9918]	−0.2695 [−12.1569]	−0.0009 [−6.6192]	−0.0014 [−8.2049]	−	0.4761	4,355
2003	0.0151 [3.4372]	−0.2741 [−11.6217]	−0.0005 [−3.0827]	−0.0017 [−7.0182]	−	0.5253	4,671
2004	0.0157 [4.2885]	−0.3125 [−15.4105]	−0.0022 [−16.8397]	−0.0060 [−20.0556]	−	0.4954	4,686
2005	0.0164 [1.7226]	−0.3054 [−11.1816]	−0.0021 [−7.9847]	−0.0021 [−3.5517]	−	0.4660	4,530
2006	0.0291 [2.9897]	−0.3037 [−10.5964]	−0.0020 [−9.9534]	−0.0067 [−8.4785]	−0.1131 [−8.2678]	0.4151	3,557
2007	0.0383 [4.4645]	−0.2202 [−12.2163]	−0.0023 [−39.6644]	0.0013 [4.3237]	−0.1096 [−25.3770]	0.4026	2,566
2008	0.0146 [1.2308]	−0.0787 [−2.7220]	0.0025 [104.0220]	0.0062 [14.4361]	−0.0652 [−11.3025]	0.5753	2,116
2009	−0.0038 [−0.7136]	−0.1607 [−6.3107]	−0.0005 [−6.8303]	−0.0002 [−1.2805]	−0.0809 [−8.1676]	0.6897	2,696
2010	0.0081 [1.4773]	−0.1387 [−7.0479]	−0.0001 [−0.6642]	−0.0019 [−10.2903]	−0.0879 [−15.9156]	0.6681	3,057
2011	0.0029 [0.4408]	−0.1701 [−6.4558]	0.0055 [41.3848]	0.0046 [39.8818]	−0.0787 [−12.7182]	0.6612	3,088
2012	0.0022 [0.2513]	−0.1730 [−6.8584]	−0.0025 [−13.3691]	−0.0050 [−32.4328]	−0.1083 [−13.2538]	0.5881	4,176

Table 4: Coefficient estimates for year by year regressions based on equation 2 for the city of San Diego. T-stats are reported in brackets and are based on standard errors clustered by the year the transaction was initiated and by census tract.

	$IsRemodeled$	$\log(P)$	$\log(P) \cdot I_{3y:6y}$	$\log(P) \cdot I_{>6y}$	$IsShort$	R2	N Obs
2001	0.0094 [2.2697]	−0.2809 [−23.0400]	−0.0012 [−14.4510]	0.0004 [1.2238]	−	0.4181	7,313
2002	0.0173 [5.2578]	−0.2666 [−23.9733]	−0.0017 [−21.0963]	−0.0022 [−25.4238]	−	0.5215	10,299
2003	0.0165 [3.7183]	−0.3169 [−20.7563]	−0.0024 [−12.4727]	−0.0037 [−14.7576]	−	0.6170	11,255
2004	0.0172 [5.5196]	−0.3121 [−18.8920]	0.0010 [2.7695]	−0.0013 [−2.9027]	−	0.5880	12,406
2005	0.0099 [2.7855]	−0.3376 [−13.7549]	−0.0028 [−9.2372]	−0.0056 [−9.5575]	−	0.6233	11,831
2006	−0.0001 [−0.0177]	−0.3662 [−17.6542]	−0.0017 [−8.4675]	−0.0052 [−10.1448]	−0.2059 [−5.7792]	0.5516	9,187
2007	0.0227 [4.0566]	−0.3209 [−14.9379]	−0.0015 [−6.9165]	0.0006 [0.9456]	−0.1199 [−13.4030]	0.4550	6,753
2008	0.0384 [8.8965]	−0.1789 [−5.3112]	0.0031 [16.7093]	0.0059 [9.3147]	−0.0652 [−18.0938]	0.6824	5,102
2009	0.0209 [3.4501]	−0.0699 [−2.7729]	−0.0034 [−37.9278]	−0.0064 [−14.0018]	−0.0944 [−9.2579]	0.7425	5,888
2010	0.0191 [2.3809]	−0.1244 [−2.8314]	0.0011 [3.1184]	0.0008 [2.0070]	−0.1130 [−15.5214]	0.7803	6,822
2011	0.0164 [2.7578]	−0.1123 [−3.1655]	0.0041 [5.9256]	0.0029 [4.2092]	−0.1195 [−15.7044]	0.7686	6,888
2012	0.0082 [0.9047]	−0.1032 [−3.1963]	−0.0025 [−4.2061]	−0.0046 [−6.9201]	−0.1716 [−14.6658]	0.7349	8,520

Table 5: Coefficient estimates for year by year regressions based on equation 2 for the city of San Francisco. T-stats are reported in brackets and are based on standard errors clustered by the year the transaction was initiated and by census tract.

Table 6 also reports Moran's I for the local market equivalents lme_{fwd}. Local market equivalents show positive correlations, especially in the years of the housing bust. Thus, the conditioning

information used in equation 2 seems to do a good job at accounting for factors that determine differences in abnormal performance across different locations. The residuals are close to *i.i.d.* in a spatial sense and therefore a good proxy of house-specific shocks.

	Residuals Moran's I (%)			lme_{fwd} Moran's I (%)		
	Los Angeles	San Diego	San Francisco	Los Angeles	San Diego	San Francisco
2001	0.35	0.82	0.34	3.19	4.77	6.98
2002	0.44	1.43	0.55	2.92	8.03	12.01
2003	0.37	0.76	0.78	2.83	10.61	15.87
2004	0.30	0.85	1.13	4.19	11.40	14.44
2005	0.30	2.70	1.01	7.45	9.20	14.20
2006	0.44	1.25	0.83	10.06	7.53	8.23
2007	0.18	1.52	0.96	5.92	3.26	4.59
2008	0.23	0.80	0.64	22.56	25.26	23.27
2009	0.62	1.33	2.14	24.81	31.84	32.12
2010	0.69	1.05	1.93	26.74	31.30	33.93
2011	1.18	1.17	2.40	25.01	25.09	31.79
2012	0.64	1.01	1.51	21.29	18.65	21.55

Table 6: Moran's I for the residuals from regression equation 2 and for lme_{fwd} in each metropolitan area and sale year.

Figure 2 shows the distribution of residuals for each year from 2001 to 2012 for the city of Los Angeles. The empirical distributions are not skewed, but slightly leptokurtotic (see tables 16 and 17 in appendix A). Residual dispersion is sizable and increases after 2008. Year by year distributions of residuals for San Diego and San Francisco are also reported in figures B.26 and B.27 in appendix B and show similar evidence.

Figure 3 shows measures of the cross-sectional dispersion of residuals in each year. The panels in the first row report standard deviations, while the ones in the second and third row show inter-quantile differences. Cross-sectional dispersion is very large, and increases across all three metropolitan areas during and after the housing crisis. In Los Angeles, the standard deviation goes from approximately 10% in 2006 to 13% in 2012. In San Diego, it goes from slightly below 9% to 10%, while in San Francisco it goes from slightly above 10% to 13%. A similar behavior can be observed for the differences between the 75th and 25th quantile and between the 90th and 10th quantile from the panels respectively in the second and third row of figure 3.

The panels in the first row of figure 3 also report year by year standard deviations of the local market equivalents (lme_{fwd}). We can see that the increase in standard deviation during and after the crisis is substantially larger for local market equivalents rather than for the residuals. However, most of the increase is explained by house and trade characteristics and by local market fluctuations. The increase in the actual house-specific component is substantially smaller.

So far idiosyncratic risk has been computed by pooling house-specific shocks for transactions with different characteristics. In particular, by pooling resales with different holding periods. This
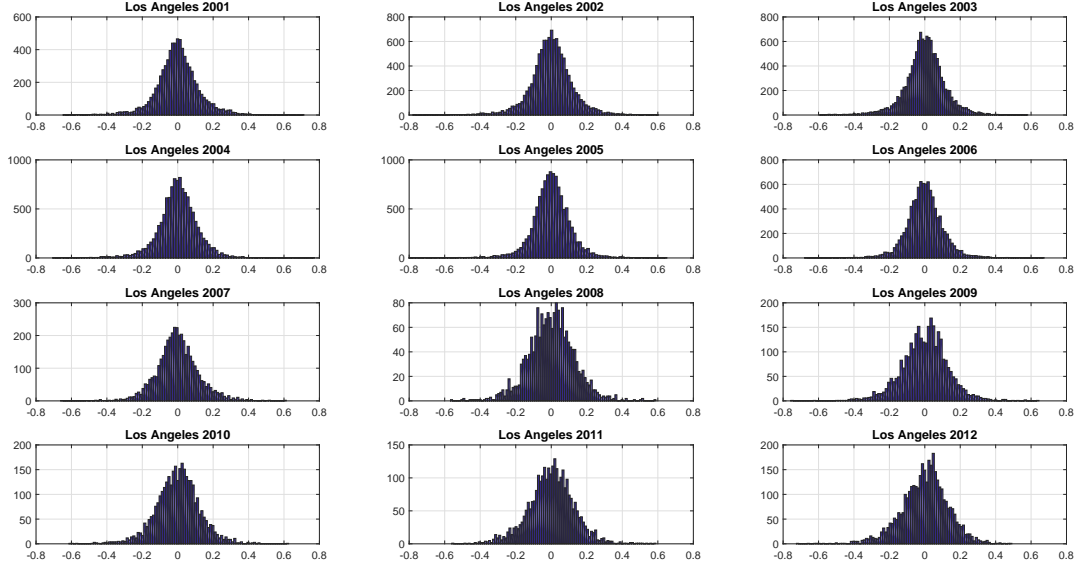
Figure 2: Distribution of residuals from regression equation 2, years from 2001 to 2012, Los Angeles.
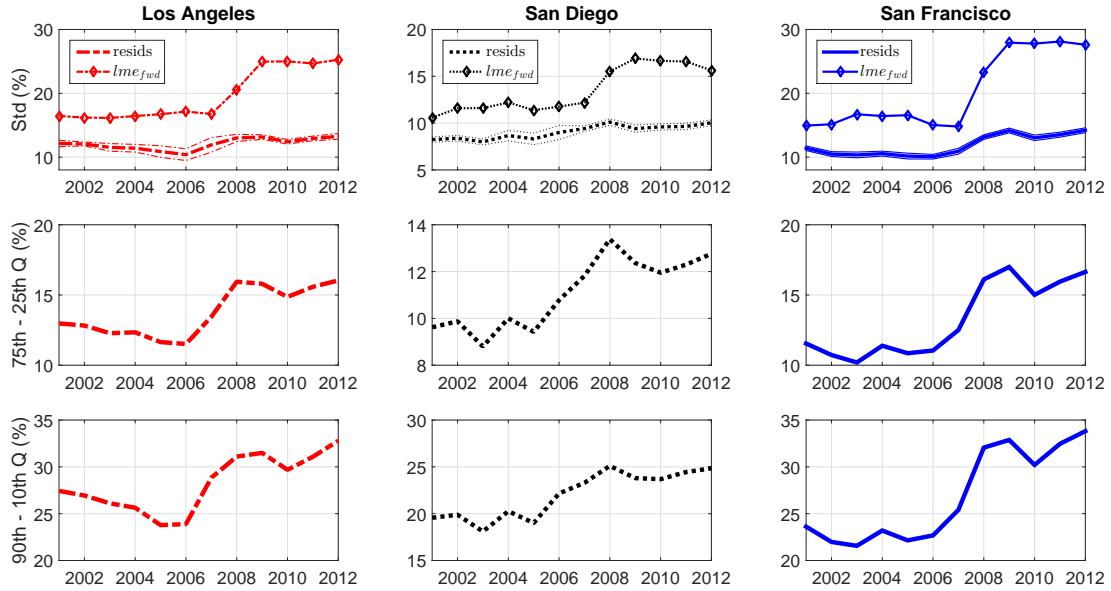

Figure 3: Measures of cross-sectional dispersion of residuals from equation 2. The panels in each column correspond to a different city, while the different cities correspond to different dispersion measures: standard deviation, inter-quartile difference and inter-decile difference. The sample covers year from 2001 to 2012.

is problematic. For example, if we believe that idiosyncratic risk is a sum of annual *i.i.d.* shocks, it would make more sense to annualize house-specific residuals. The next section is therefore devoted to further studying the properties of idiosyncratic shocks and the importance of composition effects in the measurement of idiosyncratic risk.

The preceding calculations omit rent income, which is a significant part of housing returns. A concern with my results in this section is that idiosyncratic risk might be dwarfed by the magnitude and more importantly by fluctuations in rent rates. It is therefore important to assess the relative magnitude of idiosyncratic price risk with respect to rent rates.

Information on residential real estate rents in California is provided by the US department of Housing and Urban Development [11] (HUD). The data consist of average annual fair market rents and are available at county level for houses of different "sizes", defined in terms of number of bedrooms (ranging from one to four). County level house price indexes for houses with different numbers of bedrooms are available from Zillow[12]. By merging the two datasets, I can compute rent-price ratios for each county and for houses with different numbers of bedrooms.

Figure 4 reports the evolution of the rent-price ratio for houses with 2 bedrooms[13] for the metropolitan areas of Los Angeles, San Diego and San Francisco.
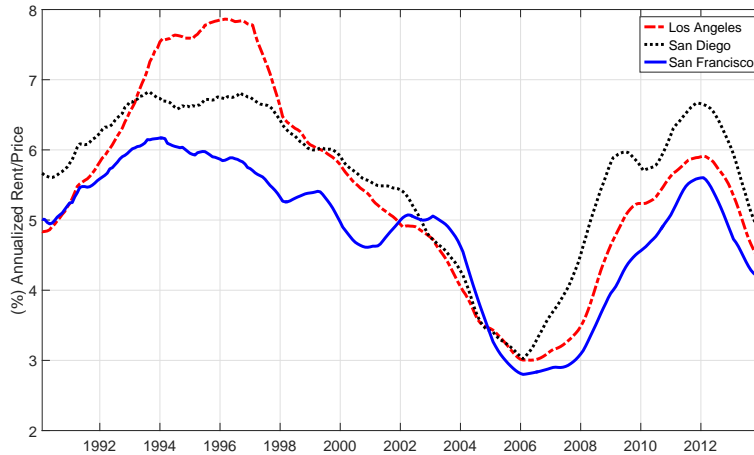

Figure 4: Annualized rent/price ratio for houses with 2 bedrooms in the metropolitan areas of Los Angeles, San Diego and San Francisco. Monthly data over the sample period between January 1990 and December 2013.

Rent rates co-move across cities, similarly to what we have seen for house price indexes in figure

---

[11] Available at: http://www.huduser.org/portal/datasets/fmr.html

[12] House price index data are available from the Zillow Home Value Index dataset (ZHVI®) and can be downloaded at http://www.zillow.com/research/data/.

[13] The behavior of the rent-price ratio for houses with 1,3 and 4 bedrooms is similar to the one reported in figure 4. The graphs for these series are available upon request.

1. Rents have been stickier than house prices in the 2000s, so that figure 4 shows counter-cyclical fluctuations of annual rent-price ratios. The series hit a bottom low in 2006 at 3 % and rise up to 5 % or 6 % (depending on the city) in 2010. Over the entire sample from January 1990 to December 2013, the average annual rent rate is close to 4.5 % for San Francisco and slightly above 5 % for Los Angeles and San Diego.

While the rent-price ratio is volatile, idiosyncratic price risk appears to be large both with respect to the fluctuations and the level of rent rates. Both in 2001 and 2010, one standard deviation of the residuals from equation 2 is from two to three times as large as the annual rent payment. This ratio becomes even larger during the housing boom.

4 The Properties of Idiosyncratic Risk

In this section I investigate factors that may determine differences in idiosyncratic risk across different house resales. A key focus is the relationship between idiosyncratic risk and home holding period, even if I also explore other dimensions of the data.

In fact, differences in idiosyncratic risk across different houses could be driven by local geographic factors, by characteristics of the trade and by characteristics of the asset that is being traded.

I start by assessing the importance of potential geographic factors. I do this by looking at spatial correlations of squared residuals. Estimated correlations are small, always close to zero (see table 7). This suggests that local geographic factors do not play a major role in determining the distribution of squared residuals.

	Los Angeles	San Diego	San Francisco
2001	1.80%	0.81%	0.62%
2002	1.69%	1.12%	1.19%
2003	1.09%	1.66%	0.72%
2004	0.80%	3.30%	0.86%
2005	1.06%	5.65%	0.73%
2006	1.71%	1.47%	0.84%
2007	1.56%	1.44%	1.61%
2008	1.03%	1.20%	1.57%
2009	1.18%	2.37%	2.76%
2010	1.45%	1.69%	1.96%
2011	0.98%	1.37%	2.53%
2012	0.93%	2.34%	1.62%

Table 7: Moran's I for squared residuals from regression equation 2 for each year and city.

House and sale characteristics may drive differences in house-specific risk. I study the behavior of residuals dispersion in subsamples split according to whether houses were sold trough a distress sale and whether houses were remodeled. Figure B.28 in appendix B compares city by city the

idiosyncratic risk for normal and short sales in each year between 2001 and 2012. Short sales standard deviation is larger for Los Angeles and San Francisco in 2006. However, short sales were very infrequent at the top of the housing boom, and the 2006 standard deviation is computed using very few observations. The dispersion of idiosyncratic residuals for short sales falls in 2007 and then co-moves with the one of normal sales from there on. If anything, the standard deviation of residuals from short sales in Los Angeles is smaller than the one for normal sales. Figure B.29 in appendix B compares the standard deviation of residuals from 2 for houses that respectively were and were not remodeled. In Los Angeles residuals standard deviation is larger for houses that were remodeled until 2007. Houses that were and were not remodeled appear to have similar standard deviation of residuals in San Diego and San Francisco.

The relation to the holding period is a key dimension of idiosyncratic risk[14]. Do house-specific shocks accumulate and scale with time? Do homeowners with longer holding periods experience larger idiosyncratic risk when selling their house? I estimate residuals variance conditional on holding period using a non-parametric estimator developed by Phillips and Xu (2011) [15], which is in principal similar to a kernel regression of squared residuals on holding period. Estimates of conditional standard deviation as a function of holding period for the cities of Los Angeles, San Diego and San Francisco and years 2001, 2004, 2008 and 2012 are reported in figures 5, 6 and 7.

---

[14]Another potential driver of differences in idiosyncratic risk is house quality. While actual house quality is not observable, I use house price as a proxy. I split house sales based on whether at the time the house was bought, its price was above or below the median price across all transactions taking place in the same metropolitan area and the same year (when the house was bought). I then compute standard deviations for the two groups in each "sale" year. Figure B.30 in appendix B shows that houses with above median prices tend to have a more stable level of idiosyncratic risk over time, and therefore have larger standard deviation during the boom years and smaller standard deviation during the bust. These differences are small in Los Angeles and San Diego, but evident in San Francisco. Idiosyncratic risk is therefore more cyclical for cheaper houses. While this finding is interesting, the effect is large only in San Francisco. I therefore postpone a more detailed analysis of this aspect of the data to future work.

[15]Conditional variance is computed as:

$$\hat{\sigma}_i^2 = \frac{\sum_{j=1}^{N_{T_i}} \hat{w}_{j,i} K_H (X_j - X_i) \hat{\varepsilon}_{fwd,j}^2}{\sum_{j=1}^{N_{T_i}} \hat{w}_{j,i} K_H (X_j - X_i)}$$

Where $\hat{\varepsilon}_{fwd,j}$ are residuals from equation 2 and the vectors X_j and X_i are row vectors containing latitude and longitude of houses j and i. I use a normal kernel $K_H(Z) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2} Z'HZ\right)$, where H is a diagonal matrix with the square of bandwiths. The weights $\hat{w}_{j,i}$ solve the empirical likelihood problem:

$$\{\hat{w}_{1,i}, ..., \hat{w}_{N_{T_i},i}\} = \underset{\{w_{1,i}, ..., w_{N_{T_i},i}\}}{\arg\min} \quad -2\sum_{j=1}^{N_{T_i}} \log\left(N_{T_i} w_{j,i}\right)$$

$$subject \; to$$
$$w_{j,i} \geq 0 \;\; \forall \; j$$
$$\sum_{j=1}^{N_{T_i}} w_{j,i} = 1$$

The 95% confidence intervals are computed by simulation using a sub-sampling algorithm.

In each figure, I add a term structure of standard deviation generated under the null of *i.i.d.* monthly shocks and calibrated to match standard deviation estimates from the data for an holding period of two years. Estimates from the data are very close across different holding periods. The "term structure" of standard deviation is far off from the one that would prevail under the null of risk scaling with time.
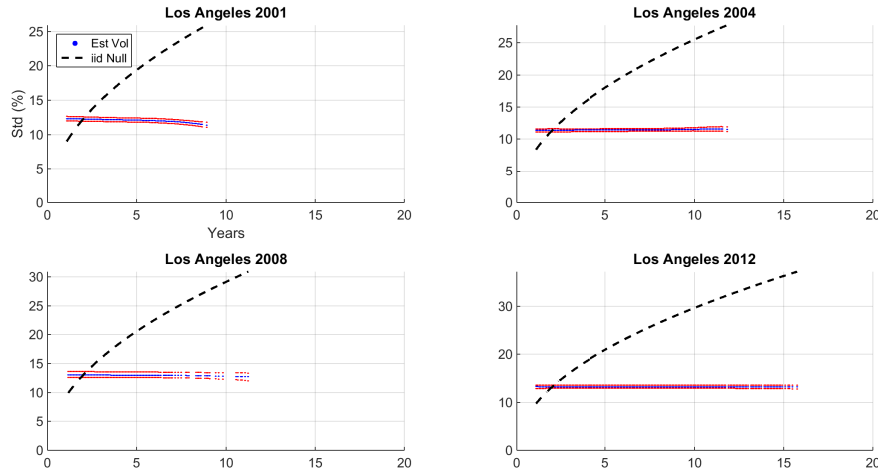

Figure 5: Conditional standard deviation estimates by holding period against null of *iid* shocks. City of Los Angeles, different panels correspond respectively to years 2001, 2004, 2008 and 2012.
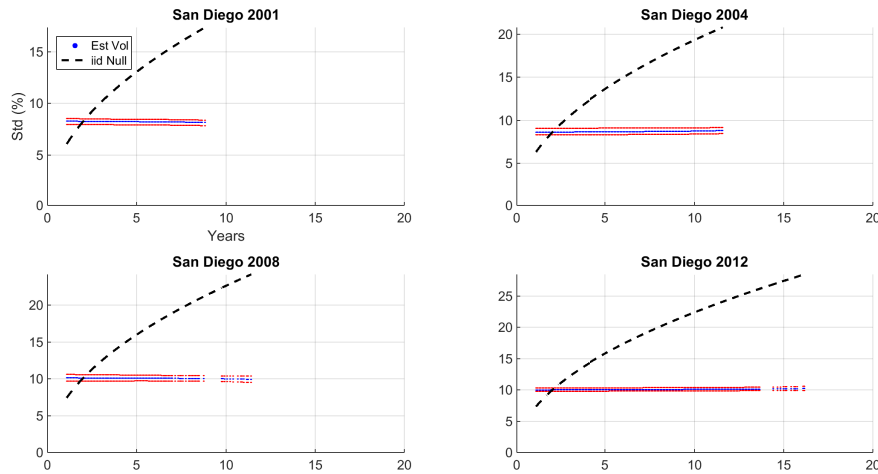
Figure 6: Conditional standard deviation estimates by holding period against null of *iid* shocks. City of San Diego, different panels correspond respectively to years 2001, 2004, 2008 and 2012.
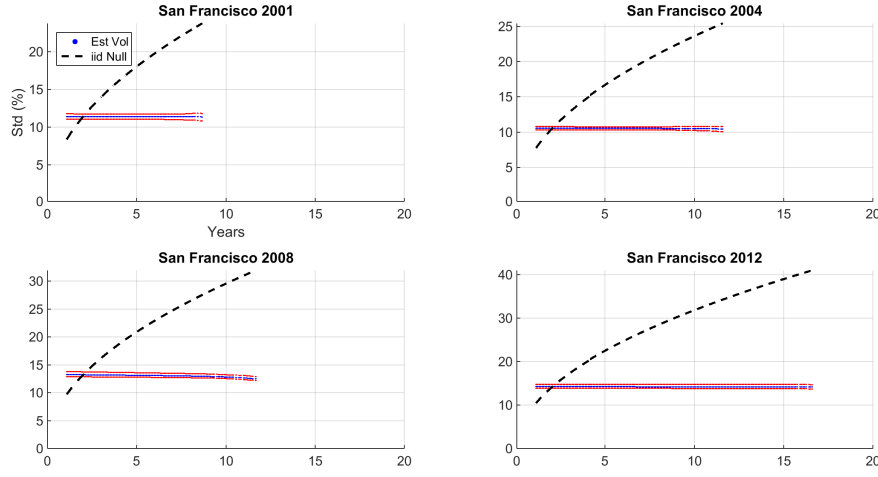
Figure 7: Conditional standard deviations estimates by holding period against null of *iid* shocks. City of San Francisco, different panels correspond respectively to years 2001, 2004, 2008 and 2012.

Figure 8 shows the evolution of house-specific risk over time for housing units held for three, five and eight years. It is again evident that there is little differences in idiosyncratic standard deviation across holding periods.
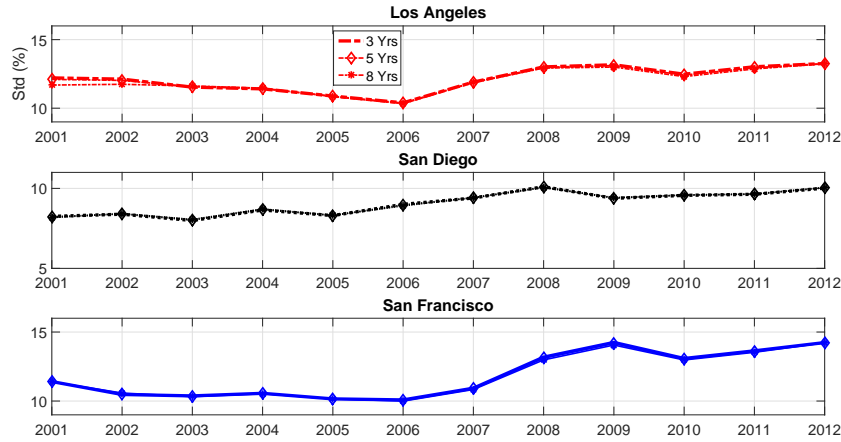

Figure 8: Time series of standard deviation estimates for houses with holding periods of three, five and eight years in Los Angeles, San Diego and San Francisco. Period from 2001 to 2012.

I now propose an alternative approach to illustrate the relationship between idiosyncratic housing risk and holding period. I provide a decomposition of the idiosyncratic component of capital gains into a transitory and a persistent shock. This methodology is widely use in the literature that studies idiosyncratic labor income risk (see for example Meghir and Pistaferri (2004), Storesletten et al. (2004) and the literature review by Meghir and Pistaferri (2011)). Consider $\varepsilon_{i,j,t}$, the abnormal

capital gain of house i, held for j years and sold in year t. I can write the decomposition of the abnormal capital gain into a transitory and a persistent shock, in formulas:

$$\varepsilon_{i,j,t} = \eta_{i,j} + \sigma_{u,t}u_{i,t} \tag{3}$$

$$\eta_{i,j} = \rho\eta_{i,j-1} + \sigma_v v_{i,j}$$

Where $\eta_{i,j}$ is the persistent component, which accumulates over the house holding period. $u_{i,t}$ is the transitory shock, which is determined at one point in time during the life of the trade. For simplicity, in 3 I assume that the transitory shock is realized at the time of sale. The coefficients $\sigma_{u,t}$, ρ and σ_v can be estimated using moments of squared abnormal capital gains. In fact, since the panel regressions in 3 contain year fixed effects, the model in 3 implies the following moment conditions:

$$g(\sigma_{u,t}, \rho, \sigma_v, j) = \sum_{i=1}^{N} \frac{1}{N_{t,j}} \varepsilon_i^2 I_{t,j} - \left(\sigma_{u,t}^2 + \sum_{k=0}^{j-1} \rho^{2k}\sigma_v^2 \right)$$

Where j is again a specific holding period, in years. N is the total number of observations in the dataset, across all holding periods and years of sale. $N_{t,j}$ is the number of observations with holding period j in year of sale t. $I_{t,j}$ is a dummy variable equal to one when an observation corresponds to a house sale in year t with holding period j. All moments across different years and holding periods are then collected in the vector $g(\chi)$, where $\chi = \{\sigma_{u,2001}, ..., \sigma_{u,2012}, \rho, \sigma_v\}$. A minimum distance estimator for the vector χ is found by solving numerically:

$$\chi^* = \min_{\chi} g(\chi)'Wg(\chi)$$

In the estimation, W is an identity matrix. This choice is based on the study by Alntonji and Segal (1996), which shows that the use of the optimal weighting matrix as suggested in Hansen and Singleton (1982) would introduce substantial small sample bias. Parameter estimates are reported in in table 8. Standard errors are in brackets and are computed using a bootstrapping algorithm. Panel (a) shows estimates of the autoregressive coefficient ρ and the conditional standard deviation of persistent shocks σ_v, as well as estimates of the standard deviation of transitory shocks $\sigma_{u,t}$. Estimates are based on the sample covering the years-of-sale from 2001 to 2012, and the coefficients are estimated separately for each city (Los Angeles, San Diego and San Francisco). Under the null that idiosyncratic risk is entirely determined by a component that scales with the investment horizon, ρ is equal to one and the standard deviation of transitory shocks is negligible. Panel (a) of table 8 describes a very different picture: ρ is very close to zero, and σ_v is very small. On the other hand, the standard deviation of transitory shocks $\sigma_{u,t}$ are large and absorb the full dispersion of the data (this is evident when comparing the estimates from the table to the statistics reported in

23

figure 3). Panels (b) and (c) of table 8 report estimates of ρ and σ_v respectively for the sub-sample containing years-of-sale from 2001 to 2006 and the one covering years-of-sale from 2007 to 2012. Again, estimates are reported separately for each city. Results are qualitatively similar to the ones obtained using all observations.

Overall, the transitory component seems to be the key driver of dispersion in abnormal capital gains, while the date reject the existence of a persistent component.

5 Robustness

My findings are robust to different methods of measuring house-specific risk. Figures 9 and 10 and 11 show standard deviation as a function of holding period for the raw local market equivalents (lme_{fwd}). Again, I compare the estimated term structure against the null of iid shocks accumulating over time. We can see that even when I am not conditioning on house and trade characteristics, the estimated term structure of house-specific risk is flat. In other words, individual house risk does not scale with holding period.
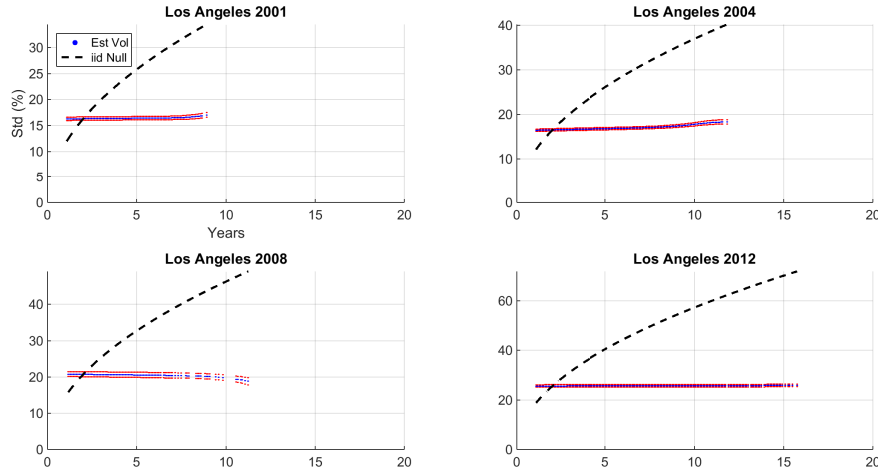

Figure 9: Conditional standard deviation estimates (based on lme_{fwd}) by holding period against null of *iid* shocks. City of Los Angeles, different panels correspond respectively to years 2001, 2004, 2008 and 2012.

As a further robustness check, I show that the evidence on the term structure of idiosyncratic risk also holds within specific subsamples of transactions. Figure 12 shows the same standard deviation estimates as in 8, but based on the sample of house resales that did not involve remodeling. Figures 13 and 14 repeat the exercise respectively for houses bought at a price above and below median. The term structure of standard deviation appears to be flat across all the different cities and years, for both cheap and expensive houses.

	Panel (a): Years from 2001 to 2012		
	Los Angeles	San Diego	San Francisco
ρ	−0.0106 (0.0082)	0.0090 (0.0059)	0.0010 (0.0071)
σ_v	0.0020 (0.0006)	0.0161 (0.0051)	0.0017 (0.0004)
$\sigma_{u,2001}$	0.1215 (0.0014)	0.0822 (0.0014)	0.1140 (0.0019)
$\sigma_{u,2002}$	0.1207 (0.0013)	0.0837 (0.0014)	0.1048 (0.0015)
$\sigma_{u,2003}$	0.1153 (0.0011)	0.0798 (0.0019)	0.1035 (0.0015)
$\sigma_{u,2004}$	0.1140 (0.0012)	0.0865 (0.0015)	0.1055 (0.0012)
$\sigma_{u,2005}$	0.1087 (0.0013)	0.0830 (0.0013)	0.1017 (0.0012)
$\sigma_{u,2006}$	0.1040 (0.0011)	0.0898 (0.0016)	0.1007 (0.0013)
$\sigma_{u,2007}$	0.1192 (0.0017)	0.0940 (0.0019)	0.1092 (0.0018)
$\sigma_{u,2008}$	0.1301 (0.0028)	0.1010 (0.0022)	0.1315 (0.0021)
$\sigma_{u,2009}$	0.1315 (0.0022)	0.0940 (0.0018)	0.1418 (0.0023)
$\sigma_{u,2010}$	0.1241 (0.0021)	0.0958 (0.0020)	0.1303 (0.0020)
$\sigma_{u,2011}$	0.1293 (0.0023)	0.0963 (0.0017)	0.1352 (0.0021)
$\sigma_{u,2012}$	0.1325 (0.0020)	0.1002 (0.0016)	0.1421 (0.0018)
$Jstat$	7.30	1.61	4.40
$Jstat$ 95th Q	12.70	4.38	8.02

	Panel (b): Years from 2001 to 2006		
	Los Angeles	San Diego	San Francisco
ρ	0.0012 (0.0080)	0.0160 (0.0078)	−0.0181 (0.0109)
σ_v	0.0021 (0.0006)	0.0070 (0.0025)	0.0025 (0.0007)
$Jstat$	4.97	0.30	2.23
$Jstat$ 95th Q	7.19	0.68	5.34

	Panel (c): Years from 2007 to 2012		
	Los Angeles	San Diego	San Francisco
ρ	−0.0131 (0.0076)	0.0054 (0.0062)	−0.0043 (0.0088)
σ_v	0.0069 (0.0045)	0.0251 (0.0073)	0.0024 (0.0007)
$Jstat$	1.86	0.90	1.73
$Jstat$ 95th Q	3.89	2.35	3.17

Table 8: Estimates, based on the micro-data, of ρ, σ_v from the system of equations 3. Panel (a) reports estimates based on data for all years-of-sale, from 2001 to 2012. The panel also reports estimates of $\sigma_{u,t}$ for each year. Panels (b) and (c) only report estimats of ρ and σ_v, respectively for the sub-sample containing years-of-sale from 2001 to 2006 and the one covering years-of-sale from 2007 to 2012. Bootstrapped standard errors are in brackets. Each panel also reports J-statistics for the objective function of the minimum distance estimator, as well as a 95% critical value. The test always fails to reject the null that the moment conditions do not hold in the data.
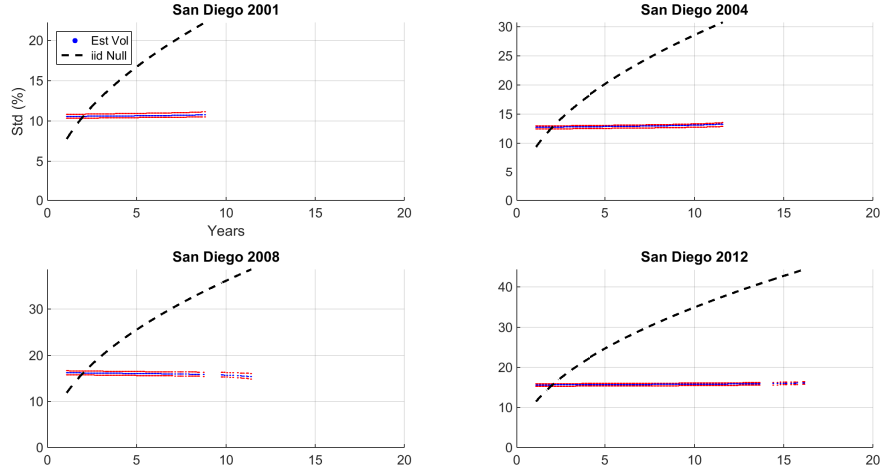
Figure 10: Conditional standard deviation estimates (based on lme_{fwd}) by holding period against null of *iid* shocks. City of San Diego, different panels correspond respectively to years 2001, 2004, 2008 and 2012.
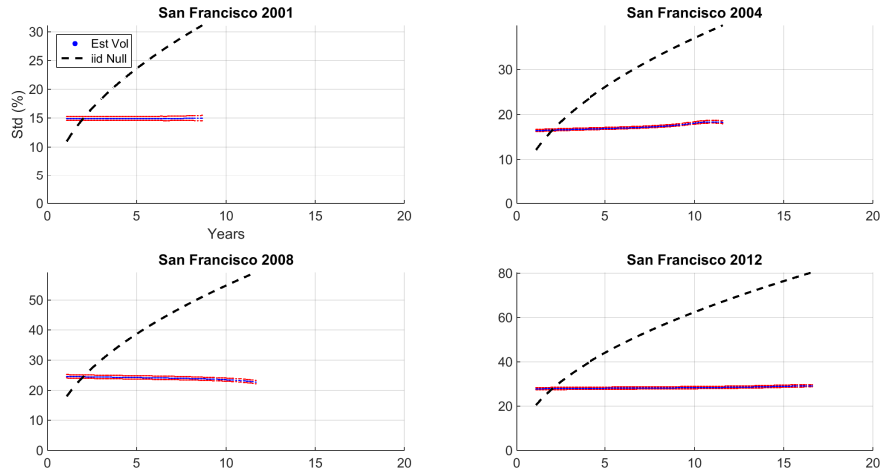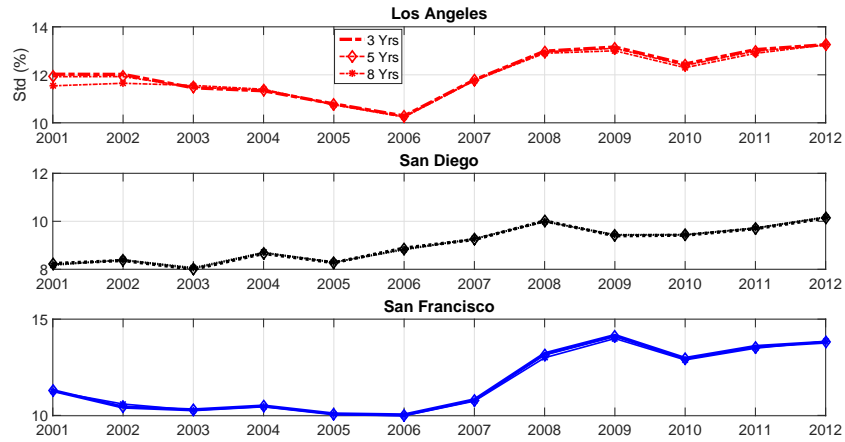

Figure 11: Conditional standard deviation estimates (based on lme_{fwd}) by holding period against null of *iid* shocks. City of San Francisco, different panels correspond respectively to years 2001, 2004, 2008 and 2012.

Figure 12: Estimates of the time series of standard deviation for houses with holding periods of three, five and eight years in Los Angeles, San Diego and San Francisco: houses that were not remodeled only. Period from 2001 to 2012.
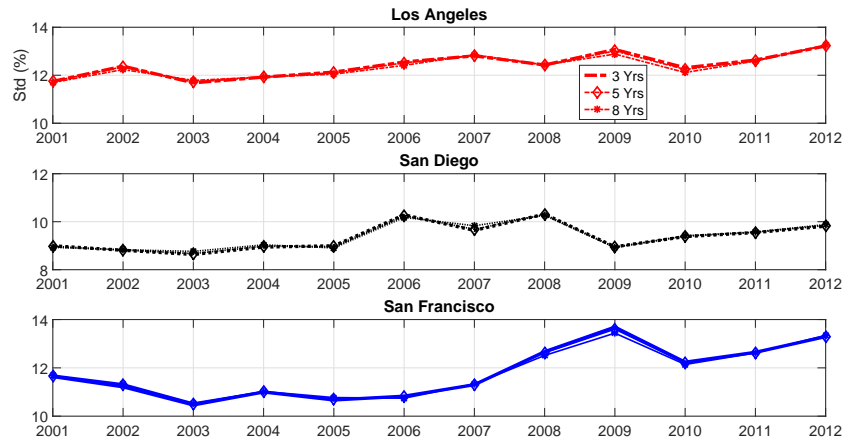

Figure 13: Estimates of the time series of standard deviation for houses with holding periods of three, five and eight years in Los Angeles, San Diego and San Francisco: houses that were bought at a price above the median transaction price of the year. Period (by sale year) from 2001 to 2012.
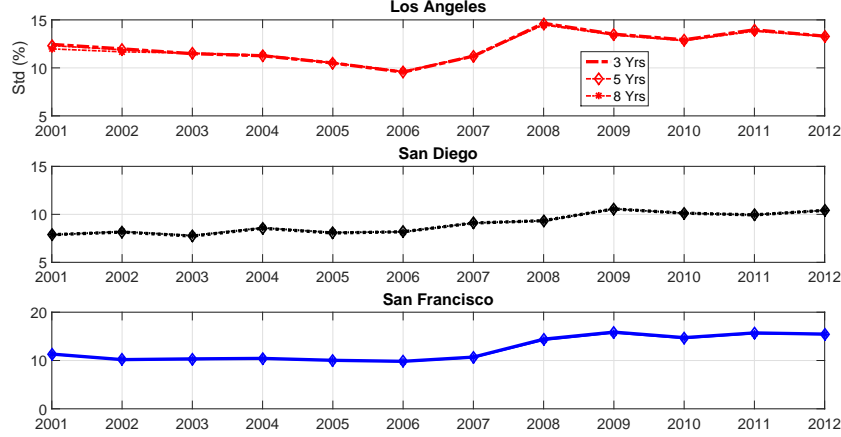
Figure 14: Estimates of the time series of standard deviation for houses with holding periods of three, five and eight years in Los Angeles, San Diego and San Francisco: houses that were bought at a price below the median transaction price of the year. Period (by sale year) from 2001 to 2012.

6 Idiosyncratic Risk for US Public Equity

The results discussed in the previous section stand in sharp contrast with evidence from other asset classes. In this section, I show that idiosyncratic risk scales with holding period for US stocks. For this new exercise, I collect data from CRSP and COMPUSTAT North America Security Monthly files. I focus on public firms which were constituents of the S&P 500 and were listed on either NYSE, AMEX or NASDAQ during the period between January 1980 and December 2013. Since stock prices are listed every day, there is no need to run hedonic regressions equivalent to equation 2. Instead, I compute risk exposures for each stock k using the five factor model from Fama and French (2014):

$$XR_{k,t+1} = a_k + b_{mkt,k}XR_{mkt,t+1} + b_{s,k}SMB_{t+1} + b_hHML_{t+1} + b_{r,k}RMW_{t+1} + b_{c,k}CMA_{t+1} + e_{k,t+1}$$

$XR_{k,t+1}$ is the capital gain of stock k in excess of the risk free rate (which is measured as the yield on one month Treasury bill) in between month t and $t+1$. $XR_{mkt,t+1}$ is the excess return of the market portfolio with respect to the risk free, SMB is the small minus big factor and HML is the high minus low book to market factor. The factors RMW and CMA are respectively the spread in returns between portfolios of stocks with robust and weak profitability and in between portfolio of stocks from firms with high and low investments (or rather pursuing conservative and aggressive investment strategies). The pricing factors as well as the one month risk free rate are

downloaded from Kenneth French's website[16].

Using coefficient estimates from the equation above, I compute the capital gains of a mimicking portfolio invested in the five factors for each stock. I then calculate idiosyncratic shocks as the difference between the terminal values of one dollar investments respectively in the stock and in the mimicking portfolio, along the same lines as in equation 1. Idiosyncratic risk is estimated as the cross-sectional standard deviation of idiosyncratic shocks, following the same methodology used in the previous sections. In particular, I compute the standard deviation of idiosyncratic shocks attained by selling stocks and mimicking portfolios on December of each year from 2001 to 2012 and across different holding periods.

Figure 15 compares, in the same fashion as figures 5, 6 and 7, the estimated "term structure" of idiosyncratic risk against a term structure that scales with holding period. This second term structure is again calibrated to match estimated idiosyncratic standard deviation at a five years holding period. In stark contrast with what I find for housing, figure 15 shows that idiosyncratic risk for stocks scales with time. Estimates from the data are very close to the ones we would obtain under the i.i.d. null[17]. This seems to be consistent with theoretical research by Cao et al. (2006), which argues that idiosyncratic equity risk is driven by idiosyncratic growth opportunities becoming available to managers over time.

7   Term Structure of Housing Risk

The fact that idiosyncratic risk does not scale with time has important implications for the composition of housing risk across different holding periods. In this section, I show that the idiosyncratic component is the dominant source of risk for short holding periods. However, on longer holding periods metropolitan-area fluctuations become the most important drivers of risk. The dynamics of metropolitan-area house price indexes are estimated using a simple time series model, described below in section 9.7.

7.1   Metropolitan Area Price Dynamics

In order to model the evolution of house price indexes, I use a simple one-lag vector autoregressive (VAR) model. The model is estimated for each metropolitan area (Los Angeles, San Diego and San

---

[16]Available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html .

[17]The magnitude of idiosyncratic stock risk from figure 15 is consistent with estimates reported by Campbell et al. (2001).
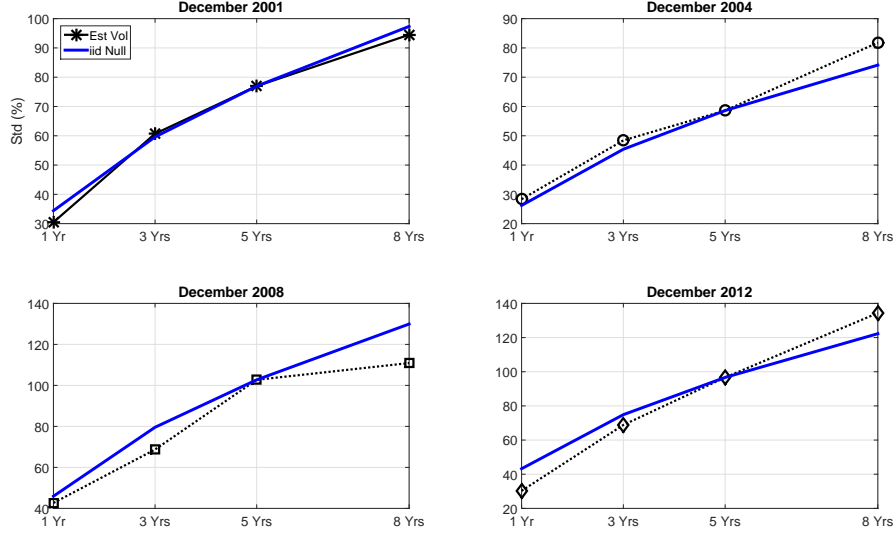
Figure 15: Idiosyncratic standard deviation for holding periods of one, three, five and eight year holding period for stocks sold in 2001, 2004, 2008 and 2012. Each panel compares standard deviation estimates based on the data against standard deviation computed under the iid null, calibrated in order to match estimates from the data at the five years horizon.

Francisco), using monthly data over the period from January 1990 to December 2013:

$$\Delta Z_{t+h} = \Gamma_0 + \Gamma_1 Z_t + \Omega v_{t+h} \tag{4}$$

The predictive horizon h is set equal to 3 months, while the elements of the vector Z_t are:

$$Z_t = [\log(P_t^{loc,r}), \log(RP_t^{loc}), inf_t^{1y}, r_t^{30y,M}]'$$

$P_t^{loc,r}$ is the real S&P Case-Shiller house price index for each metropolitan area (the time series for these indexes are reported in figure 1), while and RP_t^{loc} is the rent-price index for a two bedroom apartment (see figure 4). The variable inf_t^{1y} is twelve months inflation on all goods except food and energy from the Bureau of Labor Statistics. Finally, $r_t^{30y,M}$ is the real rate on a 30 year fixed rate mortgage. The nominal rate for 30 year US fixed rate mortgages is obtained from the St. Louis Fed database[18], to obtain the real rate I subtract one year expected inflation published by the Federal Reserve Bank of Cleveland[19]. Note that the former two time series are city specific, while the latter two are not. Table 9 shows estimates of the elements of the mean reversion matrix Γ_1, for Los Angeles, San Francisco and San Diego, along with 95 % bootstrapped confidence intervals (in

---

[18]The time series can be downloaded at https://research.stlouisfed.org/

[19]The data can be downloaded at https://www.clevelandfed.org/our-research/indicators-and-data/inflation-expectations.aspx

square brackets). The table also reports 95 % bootstrapped confidence intervals for the R-squares of the four VAR equations. Conditional moments of log price gains for metropolitan-area indexes over different predictive horizons can be computed recursively using the VAR:

$$\widehat{\Delta p_{t+k}} = E\left[\Delta p_{t+k}|Z_t\right] \quad = \Gamma_0 + \Gamma_1 E\left[\Delta p_{t+k-1}|Z_t\right]$$
$$\sigma^2_{\Delta p,k} = Var\left[\Delta p_{t+k}|Z_t\right] = \Gamma_1 Var\left[\Delta p_{t+k-1}|Z_t\right]\Gamma'_1 + \Omega\Omega'$$

Where k is the predictive horizon, in this case expressed in quarters (since the VAR has quarterly lags).

| | **Los Angeles** | | | |
	$\log(P)$	$\log(RP)$	inf^{1y}	$r^{30y,M}$
$\log(P)$	-0.1865 [$-0.2654,-0.1384$]	-0.1890 [$-0.2719,-0.1328$]	-0.7313 [$-1.3169,-0.1814$]	-1.4910 [$-2.1108,-0.8636$]
$\log(RP)$	0.1672 [$0.1078,0.2217$]	0.1578 [$0.0843,0.2120$]	0.5327 [$0.0789,1.1069$]	1.5361 [$0.9672,2.1087$]
inf^{1y}	0.0052 [$0.0015,0.0120$]	0.0045 [$0.0003,0.0120$]	-0.1080 [$-0.1704,-0.0778$]	0.0974 [$0.0488,0.1516$]
$r^{30y,M}$	-0.0181 [$-0.0280,-0.0092$]	-0.0176 [$-0.0292,-0.0080$]	0.1986 [$0.1243,0.2937$]	-0.3659 [$-0.4871,-0.3009$]
R2	$(11.87\%,24.77\%)$	$(13.86\%,20.10\%)$	$(12.28\%,29.62\%)$	$(10.97\%,31.89\%)$

| | **San Diego** | | | |
	$\log(P)$	$\log(RP)$	inf^{1y}	$r^{30y,M}$
$\log(P)$	-0.0965 [$-0.1574,-0.0478$]	-0.1108 [$-0.1742,-0.0391$]	-1.3174 [$-1.9018,-0.8180$]	-0.4404 [$-1.0475,0.1807$]
$\log(RP)$	0.0885 [$0.0247,0.1374$]	0.0885 [$-0.0038,0.1375$]	0.6410 [$0.1571,1.2586$]	0.9887 [$0.3662,1.5876$]
inf^{1y}	0.0016 [$-0.0014,0.0080$]	0.0008 [$-0.0033,0.0082$]	-0.0896 [$-0.1549,-0.0608$]	0.0688 [$0.0220,0.1205$]
$r^{30y,M}$	-0.0157 [$-0.0257,-0.0095$]	-0.0187 [$-0.0319,-0.0120$]	0.1603 [$0.0968,0.2531$]	-0.3620 [$-0.4750,-0.2987$]
R2	$(10.72\%,21.20\%)$	$(10.26\%,23.17\%)$	$(7.66\%,21.74\%)$	$(9.77\%,19.09\%)$

| | **San Francisco** | | | |
	$\log(P)$	$\log(RP)$	inf^{1y}	$r^{30y,M}$
$\log(P)$	0.0535 [$0.0095,0.0903$]	0.1089 [$0.0641,0.1658$]	-1.0070 [$-1.7082,-0.3907$]	-0.1530 [$-0.7712,0.4375$]
$\log(RP)$	-0.0713 [$-0.1213,-0.0484$]	-0.0975 [$-0.1673,-0.0711$]	0.1859 [$-0.4462,0.7647$]	0.4529 [$-0.0840,1.0062$]
inf^{1y}	0.0014 [$-0.0014,0.0046$]	0.0006 [$-0.0031,0.0044$]	-0.0854 [$-0.1499,-0.0540$]	0.0593 [$0.0211,0.1079$]
$r^{30y,M}$	-0.0016 [$-0.0084,0.0030$]	-0.0004 [$-0.0085,0.0059$]	0.1301 [$0.0461,0.2294$]	-0.2312 [$-0.3446,-0.1819$]
R2	$(12.41\%,16.40\%)$	$(12.27\%,15.24\%)$	$(8.88\%,10.70\%)$	$(11.02\%,16.25\%)$

Table 9: Estimates for the elements of Γ_1 from equation 4, sample from January 1989 to December 2014. Bootstrapped 95 % confidence intervals are reported in square brackets. I report 95 % bootstrapped confidence intervals for the R2 of each one of the four equations in the model.

7.2 Metropolitan-Area and Idiosyncratic Risk by Holding Period

I now consider an individual housing unit i, whose price risk is determined by the standard deviation of the house price index and the standard deviation of idiosyncratic shock. This is a house with a "beta of one" with respect to the house price index. Since the idiosyncratic shock is orthogonal to index fluctuations, I can write the total house price standard deviation over horizon k as:

$$\sigma_{k,i} = \sqrt{\sigma_{\Delta p,k}^2 + \sigma_\varepsilon^2}$$

Where σ_ε^2 is the variance of idiosyncratic shocks. Figure 16 show the decomposition of capital gains standard deviation into a "metro area" and an idiosyncratic component, across investment horizons from one to ten years[20]. Results are reported separately for each metropolitan area.

The standard deviation of idiosyncratic shocks matches estimates from each metro-area in year 2001. We can see that idiosyncratic risk has the highest impact for short holding periods. After one year, aggregate fluctuations account for only approximately 50 % of total standard deviation in Los Angeles and San Francisco, and approximately 60 % in San Diego. As the horizon increases the importance of the idiosyncratic shock goes down, even though at the 10 years horizon it still accounts for around 10 % of standard deviation in Los Angeles and San Francisco and 5 % in San Diego.
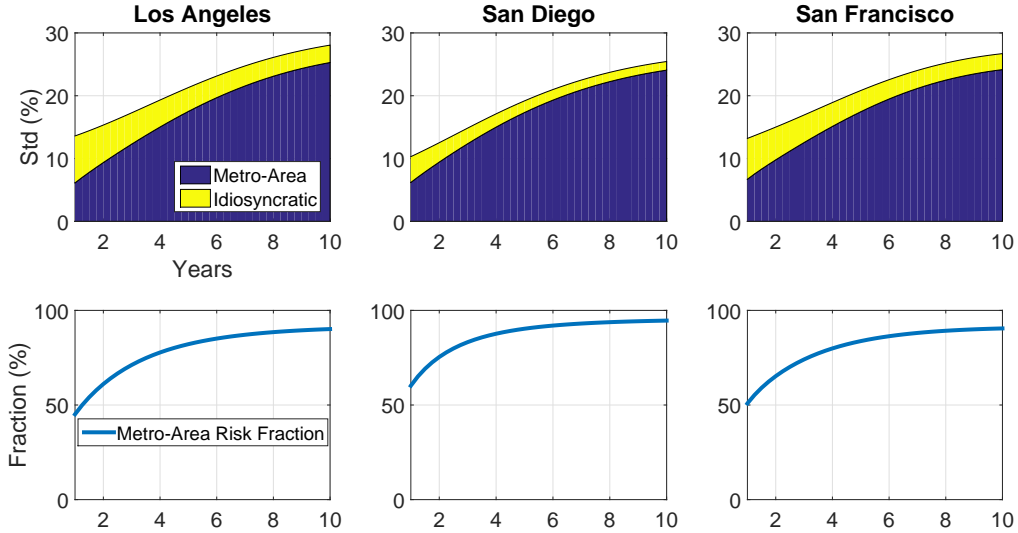

Figure 16: Composition of individual house risk across holding periods from one to ten years. Results are based on 2001 estimates.

---

[20]I do not consider horizons shorter than one year since, as explained in section 2, I exclude from the calculations based on the micro data all house sales with holding period shorter than one year.

8 Quantitative Model

I now introduce the quantitative portfolio model used to asses the cost of idiosyncratic risk from the perspective of a homeowner. Preferences are defined over a bundle of numeraire good and housing services. The homeowner may sell her house over a finite horizon T either to optimally re-balance her consumption bundle and asset portfolio or due to an exogenous mobility shock. The welfare cost is measured as the insurance premium that the homeowner would be willing to pay against idiosyncratic risk. In particular, I measure the premium that the household would pay over horizon T to remove idiosyncratic risk at the time of selling her house.

A key parameter in the model is the probability that a household may face an exogenous mobility shocks. The first part of the paper shows that the relative importance of idiosyncratic shock for housing risk decreases with the holding period. Households more exposed to mobility shocks will also be in relative terms more exposed to idiosyncratic risk.

The model falls in the same category as the ones developed by Campbell and Cocco (2003), Cocco (2005b) and Chen et al. (2013), where the dynamics of prices, interest rates, rent rates and income shocks are exogenously determined. The following section focuses on the structure of the model and on how the cost of idiosyncratic risk is assessed. Then, section 9 explains how the model is calibrated to data and section 10 discusses the quantitative results.

8.1 Preferences

The household has preferences over two goods: the numeraire consumption good c and housing services h. These goods are consumed in the bundle:

$$C = c^{1-\delta} h^{\delta}$$

The Cobb-Douglas specification of the consumption bundle fits empirical evidence on consumer behavior, as I explain in detail in section 9. The coefficient δ captures the elasticity of substitution between numeraire consumption and housing services. Wealth that is not consumed can be invested in a bank account that pays a fixed real rate per period. The household is a homeowner in the initial period t_0. Time is finite, the planning horizon consists of T periods, which correspond to T years. In the final period the household receives terminal utility λC_{t_0+T}, with $\lambda < 1$. The λ coefficient introduces in reduced form a "bequest" motive. I use this parameter to obtain from the model a pattern of numeraire consumption and non-housing wealth accumulation that is consistent with the data. More details on the role of λ are provided in section 9.9. The household has recursive preferences over the stochastic stream of consumption bundles $\{C_{t_0}, C_{t_0+1}, ..., C_{t_0+T}\}$, as in Epstein

and Zin (1989):

$$V_t = \left\{ C_t^\psi + \beta \mathbb{E} \left[V_{t+1}^{1-\gamma} \right]^{\frac{\psi}{1-\gamma}} \right\}^{\frac{1}{\psi}}$$

With $\psi = 1 - 1/\sigma$. These preferences disentangle the coefficient of relative risk aversion γ from the coefficient of inter-temporal substitution (IES), given by σ. Power utility is obtained as a special case when $\gamma = \frac{1}{\sigma}$.

8.2 House Sale and Welfare Cost

The household is a homeowner at time t_0. She may move once in between t_0 and $t_0 + T$, for following an endogenous or an exogenous shock. If the household sells her house at time τ, with $t_0 < \tau < t_0 + T$, she receives the house at price:

$$P_{i,\tau}^r = P_{i,t_0}^r R_{t_0,\tau}^{loc,r} e^{u_{idio}} = P_{i,\tau}^{loc,r} e^{u_{idio}} \tag{5}$$

Where $P_{i,t_0}^{loc,r}$ is the house price at time t_0 in terms of January 2000 dollars, while $R_{t_0,\tau}^{loc,r}$ is the real return of the corresponding metropolitan-area (either Los Angeles, San Diego or San Francisco) house price index between t_0 and τ. The variable u_{idio} is the idiosyncratic price shock.

Except for the idiosyncratic shock, house capital gains are perfectly correlated with fluctuations in the house price index. This is a simplification, that omits the modeling of the conditional mean of local market equivalents as estimated in equation 2. In the calibration, I pick the initial house price and housing services so that they reflect a "median" house within a metropolitan area at time t_0. This "median home" can to some extent be thought as a representative housing unit whose price fully co-moves with the metropolitan-area index.

When selling her home, the household buys a new house. After the sale occurs, the household cannot sell the new house and the probability of exogenous mobility shocks is set to zero. In other words, the households lives into this house till the final year $t_0 + T$. The new house can be more or less expensive than the original home (and deliver higher or lower housing services). However, the price of the new house is perfectly correlate with fluctuations in the metropolitan area index, so that the value of the new home in the final period $t_0 + T$ is:

$$P_{i,t_0+T}^{new,r} = P_{i,\tau}^{new,r} R_{\tau,t_0+T}^{loc,r}$$

With $P_{i,\tau}^{new,r}$ equal to the price at which the new house was bought and $R_{\tau,t_0+T}^{loc,r}$ equal to the return on the metropolitan-area index between τ and $t_0 + T$.

In the first baseline version of model, which is presented in section 8.3, the household has no access to mortgage financing. Thus, she always owns the entire capital invested in her house. I

introduce debt and default in section 8.4.

In modeling u_{idio}, I consider two different cases. In the first one, the distribution of idiosyncratic house price shocks matches the one in the data. In the second one, u_{idio} is set equal to zero, so that housing risk is only determined by index (or "systemic") fluctuations. I evaluate the cost of idiosyncratic risk by comparing the time t_0 present value of expected utility in a calibration where the distribution of u_{idio} matches the data against the same quantity in a calibration where u_{idio} is set to zero. I then compute the utility cost of idiosyncratic risk as[21]:

$$cost^{idio} = 1 - \frac{V_{t_0}}{V_{t_0}^*} \tag{6}$$

The value function with the star superfix is the one obtained when u_{idio} is set to zero. While this exercise provides a first insight into the impact of idiosyncratic housing risk on household welfare, the percentage decrease in expected utility is not an immediately interpretable economic quantity.

Thus, I convert the welfare cost into an annual dollar payment. I consider the calibration where idiosyncratic house risk is set to zero. I introduce an annual payment out of household income, which I call ΔY and that will take place in each year from t_0 to $t_0 + T - 1$. For $\Delta Y = 0$, the present value of expected utility at time t_0 is equal to $V_{t_0}^*$ from equation 6. I then look for the annual payment $\overline{\Delta Y}$ determining a decrease in expected utility that matches the one due to idiosyncratic risk. In formulas, I find $\overline{\Delta Y}$ such that:

$$cost^{idio} = 1 - \frac{V_{t_0}^{*,\overline{\Delta Y}}}{V_{t_0}^*} \tag{7}$$

With $cost^{idio}$ equal to the percentage cost computed in 6. The payment $\overline{\Delta Y}$ can then interpreted as the "annual" cost of idiosyncratic risk. More precisely, this is the payment that the household would be willing to make in order to remove idiosyncratic house risk in case she sells her house over horizon T. Utility costs and insurance premiums computed from the calibrated model for different initial years t_0, different household ages and for different metropolitan areas are reported and discussed in section 10.

8.3   Baseline Model

I first introduce a baseline version of the model where there is no mortgage. The basic model is then extended in the next section. As already mentioned, in the model the homeowner may move for both endogenous and exogenous reasons. For what concerns the latter, there is a mobility shock

---

[21]This is the similar to calculations of utility cost of risk implemented by Epstein et al. (2014), even if in a different context and model.

that may occur at the beginning of each period with probability π_{mob}. The value function for the household that has not yet moved can then be written as a weighted average of the value functions for households that respectively experience and do not experience the mobility shock. The weights are determined by π_{mob}; in formulas:

$$V_{own}(X, \theta, t) = \pi_{mob} V_{mob}(X, \theta, t) + (1 - \pi_{mob}) V_{nomob}(X, \theta, t) \tag{8}$$

Where X is a vector of exogenous state variables, θ is non-housing wealth and t is the current period. The value function $V_{H,mob}(X, \theta, t)$ corresponds to a owner who is hit by the exogenous mobility shock. The opposite is true for $V_{nomob}(X, \theta, t)$. The optimization problem for the owner who is not hit by the mobility shock is:

$$V_{nomob}(X, \theta, t) = \max \begin{cases} \max_c \left\{ (1 - \beta) C^\psi + \beta f_x \left(V_{own}(X', \theta', t') \right) \right\}^{\frac{1}{\psi}} \\ V_{mob}(X, \theta, t) \end{cases}$$

subject to

$$\begin{aligned}
\theta' &= y(X) + (1 + r_b)\theta - c \\
C &= c^{1-\delta} h^\delta \\
h &= rent(t_0) \\
t' &= t + 1
\end{aligned} \tag{9}$$

Where $V_{own}(., ., .)$ is calculated as shown in equation 8. Note that I use the conditional expectation function $f_x(V) = \left\{ \mathbb{E}_x \left[V^{1-\gamma} | X_t \right]^{\frac{\psi}{1-\gamma}} \right\}$. Non-housing wealth θ is invested in a risk free bank account paying the fixed rate r_b, while annual household income is $y(.)$, dependent on the exogenous states in X. Housing services are set equal to rental income earned by the house at time t_0, and are kept constant over the entire holding period of the house.

On the other hand, when the mobility shock hits, the homeowner value function solves:

$$V_{mob}(X, \theta, t) = \max_c \left\{ (1 - \beta) C^\psi + \beta f_{x,u} \left(V_S(X', \theta', u', t') \right) \right\}^{\frac{1}{\psi}}$$

subject to

$$\begin{aligned}
\theta' &= y(X) + (1 + r_b)\theta - c \\
C &= c^{1-\delta} h^\delta \\
h &= rent(t_0) \\
t' &= t + 1
\end{aligned} \tag{10}$$

Where I use the conditional expectation function $f_{x,u}(V) = \mathbb{E}_{x,u}\left\{[V^{1-\gamma}|X_t]^{\frac{\psi}{1-\gamma}}\right\}$, where again X is the vector of exogenous state variables and u is the idiosyncratic house price shock. $V_S(.,.,.)$ is the value function at the time of sale: the household hit by a mobility shock has to sell her house in the next period. At that point, the optimization problem becomes:

$$V_S(X, \theta, u, t) = \max_{c, S_h}\left\{(1-\beta)C^{\psi} + \beta f_x(V_M(X', \theta', S'_h, t'))\right\}^{\frac{1}{\psi}}$$

$$\text{subject to}$$

$$\begin{aligned}
\theta' &= y(X) + (1+r_b)\theta - c + P(X)e^u - P(X, S'_h) \\
C &= c^{1-\delta}h^{\delta} \\
h &= rent(t_0)S_h \\
S'_h &= S_h \\
t' &= t+1
\end{aligned} \tag{11}$$

The modeling of the house price at the time of sale is consistent with the framework in equation 5. The variable $P(.)$ represents the component that perfectly co-moves with the local house price index. The shock u accounts for idiosyncratic risk. The price of the new house is $P(X, S_h)$, which is equal to:

$$P(X, S_h) = P(X)S_h$$

The variable $S_h > 0$ is optimally chosen by the household and determines how the housing services of the new house compare to the ones of the house that is sold. After moving, the agent is not subject to mobility shocks and cannot move again. Her value function then solves:

$$V_M(X, \theta, S_h, t) = \max_c\left\{(1-\beta)C^{\psi} + \beta f_x(V_M(X', \theta', S'_h, t'))\right\}^{\frac{1}{\psi}}$$

$$\text{subject to}$$

$$\begin{aligned}
\theta' &= y(X) + (1+r_b)\theta - c \\
C &= c^{1-\delta}h^{\delta} \\
h &= rent(t_0)S_h \\
S'_h &= S_h \\
t' &= t+1
\end{aligned} \tag{12}$$

In the final period $t_0 + T$ the household is either still holding the original home, or has moved to a new house with housing services $rent(t_0)S_h$. The terminal utilities for these two alternatives are

equal to:

$$V_{own}(X, \theta, t_0 + T) = \lambda \left[y(X) + (1 + r_b)\theta + \hat{P}(X) \right]$$

$$V_M(X, \theta, S_h, t_0 + T) = \lambda \left[y(X) + (1 + r_b)\theta + \hat{P}(X, S_h) \right]$$

The parameter $\lambda < 1$ creates a bequest motive, that I use to ensure that non-housing wealth accumulation is consistent with the data. More details on this issue are provided in section 9.9.

8.4   Debt and Default

US households typically buy their home with a mix of their own funds and debt in the form of a nominal fixed rate mortgage (FRM)[22]. Thus, Iextend the model by introducing debt financing.

The rest of the framework remain the same as in the baseline model. The homeowner is subject to exogenous mobility shocks of the form presented in equation 8 and if not hit by mobility shocks can optimally decide whether to move. After moving, the owner is not allowed to move again and is not subject to mobility shocks. However, debt changes the nature of the optimization problem faced by the homeowner at each stage and therefore the value functions. First, debt payments change the budget constraints. Second, the presence of a debt contract raises the issue of modeling default. In the model, I do not allow for strategic default. The owner is forced to default when she is not able to meet her mortgage payments. Similarly to what is done by Campbell and Cocco (2011), the defaulted household is forced to become a renter and stays in that state till the final period T. Since the empirical part of this project is based on data from California housing markets, I build the model so that there is no recourse on unpaid mortgage amounts. When defaulting, the household loses her house, but gets rid of her remaining debt obligations.

The homeowner in year t_0 now has a mortgage amount M. The value function for the homeowner

---

[22]Campbell and Cocco (2003) show that 70 % of newly issued mortgages in the US over the period between 1985 and 2001 where fixed rate mortgages. Piazzesi and Schneider (2016) argue that the fixed rate mortgage is the predominant form of debt financing for US residential housing.

who is hit by a mobility shock (equation 10 from the baseline model) now solves:

$$V_{mob}(X, \theta, M, t) = \begin{cases} \max_c \left\{ (1-\beta)C^\psi + \beta f_{x,u}(V_S(X', \theta', M', u', t')) \right\}^{\frac{1}{\psi}} & if\ i_{nD} = 1 \\ \max_c \left\{ (1-\beta)C^\psi + \beta f_x(V_D(X', \theta', t')) \right\}^{\frac{1}{\psi}} & \text{otherwise} \end{cases}$$

subject to

$$\theta' = y(X) + \left(1 + r_b\right)\theta - c - i_{nD}r_M(X)M - i_{nD}I_M \tag{13}$$

$$C = c^{1-\delta}h^\delta$$

$$h = rent(t_0)$$

$$M' = M - I_M$$

$$t' = t + 1$$

With $r_M(.)$ equal to the real mortgage rate payment (bear in mind that the mortgage is a FRM in *nominal* terms). I_M is the mortgage principal payments. The dummy i_{nD} is equal to one when the household is able to meet her mortgage payments, in formulas:

$$i_{nD} = \mathbb{1}\left\{ \left(y(X) + \left(1 + r_b\right)\theta - c - r_M(X)M - I_M\right) > 0 \right\}$$

When i_{nD} is equal to zero, the household defaults. She loses her house, but she is also freed of mortgage payments. In case of default, the value function is $V_D(., ., ., .)$, which solves:

$$V_D(X, \theta, t) = \max_{c, S_h} \left\{ (1-\beta)C^\psi + \beta f(V_R(X', \theta', t')) \right\}^{\frac{1}{\psi}}$$

subject to

$$\theta' = y(X) + \left(1 + r_b\right)\theta - c - h_p$$

$$h = rent(t_0)S_h \tag{14}$$

$$h_p = P(X, S'_h)e^{rp(X)}$$

$$C = c^{1-\delta}h^\delta$$

$$t' = t + 1$$

Note that the household has to make rent payments h_p each year. These payments are determined by fluctuations in metropolitan-area prices and in the ratio of rents to house prices $rp(.)$.

The value functions 9 (for the homeowner who is not hit by mobility shocks) and 12 (for the homeowner who has moved) are modified along similar lines as 13. At the time of a home sale, the

optimization problem becomes:

$$V_S\left(X,\theta,M,u,t\right) = \begin{cases} \max\limits_{c,S_h,\Delta M}\left\{\left(1-\beta\right)C^{\psi}+\beta f_x\left(V_M\left(X',\theta',M',S_h',t'\right)\right)\right\}^{\frac{1}{\psi}} & if\ i_{nD}=1 \\[2mm] \max\limits_{c}\left\{\left(1-\beta\right)C^{\psi}+\beta f_x\left(V_D\left(X',\theta',t'\right)\right)\right\}^{\frac{1}{\psi}} & \text{otherwise} \end{cases}$$

subject to

$$\begin{aligned} \theta' &= y(X)+\left(1+r_b\right)\theta - c + \\ &\quad + \left(\hat{P}\left(X\right)e^u - P\left(X,S_h\right)\right)i_{nD} - \Delta M i_{nD} \\ C &= c^{1-\delta}h^{\delta} \\ h &= rent(t_0)S_h \\ S_h' &= S_h \\ M' &= M+\Delta M \\ t' &= t+1 \\ \eta &\geq M'/P\left(X,S_h\right) \end{aligned} \tag{15}$$

The household is selling her old home and paying off the old mortgage. Moreover, she is buying the new home and optimally choosing the new mortgage amount, under the constraint that the new loan to value ratio is below a threshold η. Thus, the household is optimally choosing current numeraire consumption, the quality or size of the new house and the new amount of debt.

In the final period t_0+T, the household can be in one of three different states. Either she has defaulted, or she is still holding her original home, or she has sold her original house and is holding a new one. Terminal utilities for these three different cases are respectively:

$$V_D\left(X,\theta,t_0+T\right) = \lambda\left[y(X)+(1+r_b(X))\theta - s_D\right]$$

$$V_{own}\left(X,\theta,M,t_0+T\right) = \begin{cases} \lambda\left[y(X)+(1+r_b(X))\theta+\hat{P}\left(X\right)-M\right], & \text{if}\ i_{nD,t_0+T}=1 \\ V_D\left(X,\theta,t_0+T\right), & \text{otherwise} \end{cases}$$

$$V_M\left(X,\theta,M,S_h,t_0+T\right) = \begin{cases} \lambda\left[y(X)+(1+r_b(X))\theta+\hat{P}\left(X,S_h\right)-M\right], & \text{if}\ i_{nD,t_0+T}=1 \\ V_D\left(X,\theta,t_0+T\right), & \text{otherwise} \end{cases}$$

Where $i_{nD,T}$ is a dummy set equal to one when time T wealth, net of mortgage obligations, is greater than 0.

9 Calibration

9.1 Life-cycle Stages, Starting Period and Planning Horizon

The model is solved for households of different ages. Household age affects the income process, as I explain in detail in section 9.5. It also affects the propensity of the household to save and accumulate non-housing wealth, which driven in the model by the bequest coefficient λ. The calibration of λ is discussed in section 9.9.

Given the nature of the demographic data available, in the calibration I work with "age groups" rather than specific ages. I take the perspective of homeowners that are in between 25 and 29 years old, 35 and 39 years old and finally in between 45 and 49 years old.

The model is solved for each age group (or life-cycle stage) in each city (Los Angeles, San Diego and San Francisco). I set the initial year t_0 equal to 2001.

I set the planning horizon T equal to five years. Thus, the insurance premiums generated from the model are for a five years insurance against idiosyncratic price shocks at the time of sale.

9.2 Household Preferences

Household preferences are determined by multiple coefficients in the model. Starting from the consumption bundle over which preferences are computed, a first important parameter is δ. This is the elasticity of substitution between consumption of the numeraire good and of housing services. I set this coefficient equal to 0.2, a value that is consistent with evidence from the national income and product accounts (NIPA) tables[23], as discussed in detail by Piazzesi et al. (2007). The same 0.2 value for the elasticity of substitution is also used by Piazzesi et al. (2015) when modeling the San Diego housing market.

A second important parameter is the one that drives time preferences, β. I set it equal to 0.98, as is done by Piazzesi et al. (2007), Piazzesi and Schneider (2009a) and other financial economics research working with data from the post-second world war period.

Key parameters are the coefficients of relative risk aversion and intertemporal elasticity of substitution. In their review article, Piazzesi and Schneider (2016) propose a baseline framework with additive preferences (power utility). In fact, Piazzesi et al. (2007) and Piazzesi et al. (2015) use power utility with IES equal to 0.2, which matches empirical results in Hall (1988). Chen et al. (2013) estimate relative risk aversion and IES (along with other parameters) for recursive preferences using a structural model. They find γ to be close to 3 and σ close to 0.3 (however,

---

[23]Available from the Bureau of Economic Analysis of the Department of Commerce at http://www.bea.gov/national/nipaweb/DownSS2.asp

the estimated time preference coefficient is pretty large, leading to a β coefficient close to 0.91). Empirical work by Attanasio and Weber (1995) based on micro data from the CEX panel reports IES estimates[24] close to 0.5. Their empirical evidence is in line with results from Hansen and Singleton (1996), who point out that estimates in Hall (1988) that might be downward biased. Mindful of this empirical evidence, Piazzesi and Schneider (2009a) use recursive preferences, with relative risk aversion equal to 5 (consistent with previous papers) and IES equal to 0.5. In my model, I use this last specification and then set $\gamma = 5$ and $\sigma = 0.5$.

The last preference parameter in the model is the "bequest" coefficient λ. The value of this coefficient is calibrated using the data, once all the other parameters in the model are fixed. This will be explained in detail in section 9.9.

9.3 Mobility Shocks

The probability of exogenous mobility shocks (π_{mob}) is an important parameter in the quantitative model, due to the properties of the term structure of housing risk described in section 7. However, moving motives are hard to disentangle in the data, and exogenously determined mobility is therefore hard to measure. Nonetheless, Census data can be used to get a sense of the likelihood of moving for households of different ages.

Figure 17 shows time series for the fraction of US households who moved[25] during each year from 2000 to 2013. Panel (a) show time series for homeowners, while panel (b) focuses on renters. The figures report both the fraction of movers out of all households with household head 25 years old or older as well as break downs in four age groups. These groups correspond to household heads ages in between 25 and 29, 30 and 44, 45 and 64 and greater or equal than 65. While figure 17 reports statistics for the entire country, there is evidence that mobility rates for California match the behavior of national averages. This can be seen for example from results reported by Ferreira (2004), based on data from the Integrated Public Use Microdata Series (IPUMS).

Several interesting facts emerge from figure 17. First, homeowners are substantially less likely to move than renters. The average fraction of movers varies between 6 % and 4.5 % for owners[26], while it is approximately 25 % for renters. Moreover, the likelihood of moving is very different across age groups, and steeply decreasing in household age. In 2001, 16% of the homeowners in the

---

[24]Vissing-Jorgensen and Attanasio (2003) find values of IES above 1 when focusing on stock-holders. However, households with significant stock-holdings are a small fraction of US households.

[25]The data are collected from the Migration and Geographic Mobility tables provided by the US Census, which are based on CPS, American Community Survey, Survey of Income and Program Partecipation and Decennial Census. The tables can be downloaded at http://www.census.gov/topics/population/migration/data/tables.2014.html.

[26]Ihrke and Faber (2012) study in detail the 2010 decennial Census survey and focus on five year mover rates. They find that 22.20% of households living in owner occupied housing moved between 2005 and 2010, which is in line with a 5% annual moving rate.

youngest group (25 to 29) moved. This fraction decreases after the bust, but is still as high as 12 %
in 2013. On the other hand, only 2% of the households from the eldest group (age greater or equal
than 65) moved in each year. Similarly, more than 30 % of renters in the youngest group move
each year, while the same statistic is in between 10 % and 14 % for the oldest households.

Differences across household groups can be explained by several factors. Younger households are
more likely to reorganize their portfolio to increase their consumption of housing services (see for
example the discussion and results in Campbell and Cocco (2007)). However, young households are
also more subject to several sources of exogenous shocks: relocation due to work or family reasons
as well as changes in household composition[27].

Due to the differences in mobility across age groups, I use multiple calibrated values of π_{mob} both
across and within age groups. I first set $\pi_{mob} = 0$ for all age groups. In this case, the household is
moving only when she considers it *ex-ante* optimal. I then consider an "intermediate" value of the
probability of mobility of shocks, consistent with observed mobility for homeowners and a "high"
value of the probability of mobility shocks, closer to the fraction of movers for renters. Thus, for
households with age in between 25 and 29, I consider values of π_{mob} equal to 12% and 30%. I then
set π_{mob} equal to 10% and 25% for households with age in between 35 and 39, and equal to 5% and
20 % for households with age in between 45 and 49.

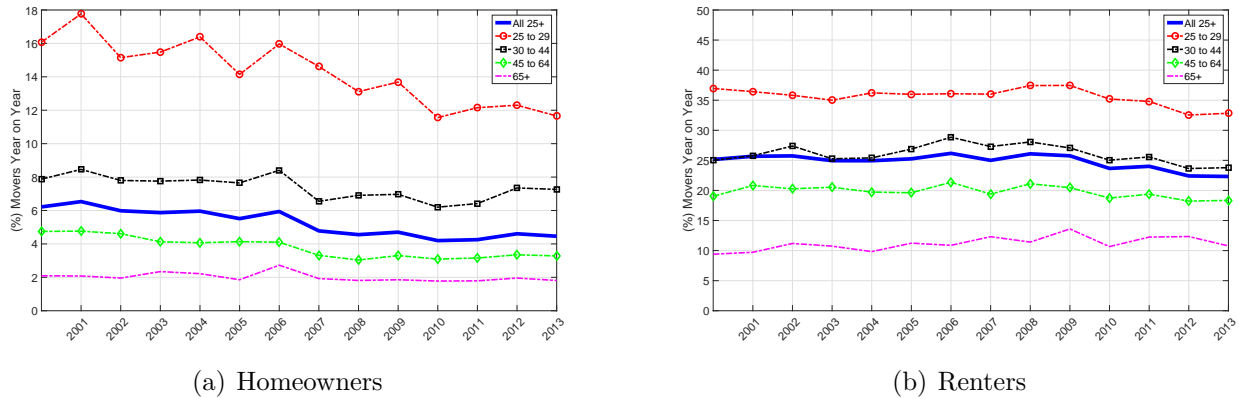(a) Homeowners   (b) Renters

Figure 17: Fraction of households belonging to different age groups who moved over the following
year. Annual data from 2000 to 2013. Based on the Migration and Geographic Mobility tables
from the US Census.

---

[27]According to data from the National Center for Health Statistics (NCHS)[28], 20% of first time marriages in the
US results in divorce or separation within 5 years. The legal procedures might lead to the sale of commonly owned
assets, such as the family house.

9.4 Household Income: Dynamics

Empirical studies in labor and macro economics decompose US households labor income in three components (see Meghir and Pistaferri (2011) for a review of the related literature). The first component is correlated with observable household characteristics. The second one is an idiosyncratic transitory shock, capturing short lived individual income fluctuations. Finally, the last component is a persistent idiosyncratic shock. In formulas, labor income for household i in year t can be written as:

$$Y_{i,t} = \tilde{Y}_{i,t} S_y(i,t) \tag{16}$$

The variable $\tilde{Y}_t$ is stochastic, while $S_y(i,t)$ is a scaling factor that adjusts income level according to observable household characteristics at time t. I discuss the calibration of this scaling factor in detail in the next section. For now, I want to focus on the idiosyncratic component of income. The log of $\tilde{Y}_t$ is $\tilde{y}_{i,t+1}$ and follows the process:

$$\tilde{y}_{i,t+1} = \rho^y_{i,t+1} + \sigma_u u_{i,t+1} \tag{17}$$

$$\rho^y_{i,t+1} = \phi_y \rho^y_{i,t} + \sigma_\eta \eta_{i,t+1}$$

The shock $\rho^y_{i,t+1}$ is persistent, with autoregressive coefficient ϕ_y, while $u_{i,t+1}$ is transitory. Both $u_{i,t+1}$ and $\eta_{i,t+1}$ are *i.i.d.* error terms. Empirical researchers working on panels of US households find that the persistent shock is either a unit root process (as in Meghir and Pistaferri (2004)), or a very slowly mean reverting shock (in Storesletten et al. (2004)). Under the unit root specification, Saporta-Eksten (2014) finds σ^2_η to be equal to 0.0225, close to the results in Meghir and Pistaferri (2004). Storesletten et al. (2004) and Cocco (2005b) estimate similar values using a mean reverting persistent process. I therefore set σ^2_η at 0.02 on annual basis, consistently with the literature. I then set ϕ_y equal to 0.98 on an annual basis, which is close to the estimates in Storesletten et al. (2004). Since in my exercise I mostly care about changes in household welfare over an horizon of multiple years, I omit from the model the transitory shock and set σ_u equal to zero.

9.5 Household Income: Life-Cycle Component

In addition to the dynamics of idiosyncratic income shocks, the other ingredient of the household labor income process is the scaling factor $S_y(i,t)$. Evidence from the labor economics literature suggests that the scaling factor is highly dependent on household age, or life-cycle stage, and household education.

I estimate the scaling factor using data from the Current Population Survey (CPS) for California. The survey provides age and education of the household head along with household income. Ideally,

I would like to use in my estimation data on household labor income net of taxes. As a first step, I compute labor income gross of taxes for each household in the survey by subtracting interest income and other non labor income from total earnings. I then get total taxes by adding up state and federal tax payments minus household tax credit. I then split households in 11 groups based on age of the household head. These groups respectively correspond to ages from 25 to 29, 30 to 34, 35 to 39, 40 to 44, 45 to 49, 50 to 54, 55 to 59, 60 to 64, 65 to 69, 70 to 74 and greater or equal than 75 years. I can further split each one of these groups based on the education attainment of the household head: high school, some college education and bachelor or higher. To estimate life-cycle component of labor income, I run the following panel regressions[29]:

$$\log(Y_i) = \alpha_t + \mathcal{B}_{age} I(age_i) + v_i \tag{18}$$
$$\log(Y_i) = \alpha_t + \mathcal{B}_{edu,age} I(edu_i, age_i) + e_i$$

The dependent variable is annual log real labor income, the coefficient α_t is an year fixed effect, $I(age_i)$ is a vector of dummy variables selecting the different age groups outlined above, while the vector $I(edu_i, age_i)$ selects educational attainment groups within each age group.

Using parameter estimates from the regressions in 18, I build estimates of the life-cycle component of income for the entire population and for subgroups based on educational attainment. These are reported respectively in the left and right hand side panels of figure 18. In particular, the left

---

[29]For each year, the survey provides sampling weights for the households in the cross-section. I call these weights $\tilde{w}_{i,t}$. I re-normalize the weights for each year t as follows:

$$w_{i,t} = \frac{\tilde{w}_{i,t}}{\sum_i \tilde{w}_{i,t}}$$

I then collect the $w_{i,t}$ in the matrix W and estimate the model above by weighted least squares with weights W:

$$\hat{\Gamma} = (X'WX)^{-1} (X'WY)$$

Where $\hat{\Gamma}$ includes the estimates of year fixed effects $\hat{\alpha}_1, ..., \hat{\alpha}_T$ and the vector of coefficients for the X variables, $\hat{\mathcal{B}}_{edu,age}$, while Z includes both the covariates X and dummies for the year fixed effects. Standard errors are obtained using a subsampling algorithm. For each draw, I normalize the weights within the draw for each year so that they sum up to one. I then estimate $\hat{\Gamma}^k$ for the kth draw. I can then compute the standard errors of estimates as:

$$SE(\alpha_t) = \sqrt{\frac{1}{K} \sum_{k=1}^{K} \left(\hat{\alpha}_t^k - \hat{\alpha}_t \right)^2}$$

$$SE(\beta_l) = \sqrt{\frac{1}{K} \sum_{k=1}^{K} \left(\hat{\beta}_t^l - \hat{\beta}_t \right)^2}$$

For each year fixed effects and each element l of β. I set $K = 1000$ draws.

hand side panel shows average log difference between the income of the group with age between 25 and 29 years old all the other "older" groups. On the right hand side, the baseline is the group with age between 25 and 29 years old and high school education. The curves reported in figure 18 are also called "Mincer curves" in the literature. The figure clearly shows how average income fluctuates over household life, increasing with seniority and then falling sharply once retirement age is reached. The pattern is similar across all education subgroups.

The Mincer curve used in the model is calibrated using estimates based on all households (the left hand side panel of 18). For each age group a, I calculate average log income growth, as the difference in log income between contiguous groups, a and $a + 1$. Then, I calculate annual average income growth g_y^a by dividing average income growth by five. This is due to the fact that age groups span five years. The scaling factor in the the model calibrated to age group a is:

$$S_y(t) = Y_{t_0} \exp\left(g_y^a(t - t_0)\right)$$

Where $t \in \{t_0, t_0 + 1, ..., T - 1\}$ is the time index keeping track of the period in the model. T is the planning horizon, which is equal to five years.
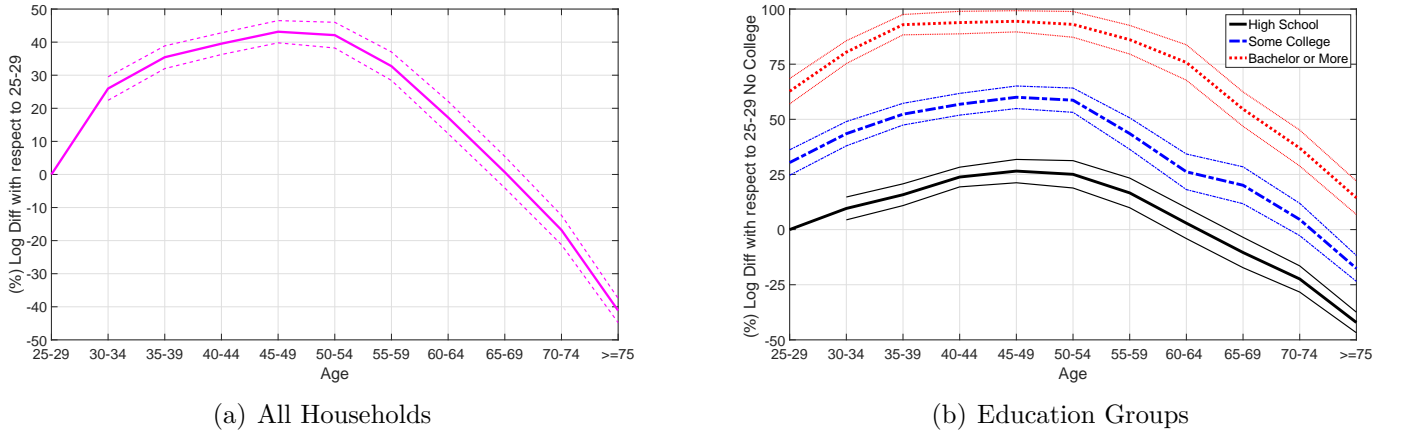

(a) All Households

(b) Education Groups

Figure 18: Mincer curves for different education groups in California. The curves are reported in terms of log differences with respect to the income of a household whose head has only attained high school education and age between 22 and 24. Estimation is based on net of taxes labor income from the Current Population Survey over the period between 1995 and 2013. The dotted lines correspond to 95% confidence intervals.

9.6 Interest Rates

The model includes two exogenously set interest rates: the rate earned on the bank account and the mortgage rate. I fix the real bank account rate to be constant and equal to 2% on an annual basis.

This first asset is risk free *in real terms*. The rate is chosen to roughly match the time preference coefficient β, which is equal to 0.98. Also, with average inflation over the period between January 1990 and December 2013 equal to approximately 2.5%, the bank account offers an average nominal rate equal to 4.5%. This is in line with the average yield on one year Treasuries over the same sample period.

The nominal mortgage rate is set to match the rate on 30 year US fixed rate mortgages at the time the mortgage is issued. This time either t_0 for the initial house or the period τ when the new house is bought. In the model, I obtain the nominal rate by summing the real mortgage rate $r^{30y,M}$ and annual inflation inf^{1y}. Since the mortgage contract is a standard nominal FRM, the real mortgage rate will fluctuate with changes in inflation. For example, in year $t > t_0$, the real mortgage rate paid by a homeowner who has not yet sold her house is computed as the real rate set at time t_0 minus the change in inflation: $r_{t_0}^{30y,M} - \left(inf_t^{1y} - inf_{t_0}^{1y} \right)$. After the homeowner sells her house, the real rate will be computed in the same way, but by taking as base year the sale period τ.

9.7 Dynamics of Exogenous States and Idiosyncratic Housing Shocks

The dynamics of the state vector X are modeled using a vector autoregression (VAR):

$$X_{t+h} = \Phi_0 + \Phi_1 X_t + \Sigma e_{t+h}^X \tag{19}$$

The horizon h is equal to 3 months. The state vector contains five variables:

$$X_t = [\log(P_t^{loc,r}), \log(RP_t^{loc}), inf_t^{1y}, r_t^{30y,M}, \rho_t^y]'$$

The first four are the elements of Z_t from equation 4. These are respectively, log real price index and log rent price index for the metropolitan area under analysis (either Los Angeles, San Diego or San Francisco), year on year inflation and real rate on a 30 year fixed rate mortgage. The last element, ρ_t^y, is the persistent shock to household income in equation 17. Parameter estimates for the dynamics of the first four variables are available from section . The dynamics of ρ^y are not estimated, but rather calibrated as explained in section 9.4. I assume that idiosyncratic income shocks have zero correlation with the other state variables in the model, so that I can write the

matrices containing VAR coefficients:

$$\Phi_0 = \begin{bmatrix} \Gamma_0 \\ \gamma_y \end{bmatrix}$$

$$\Phi_1 = \begin{bmatrix} I_4 + \Gamma_1 & 0_{3\times1} \\ 0_{1\times3} & \phi_y \end{bmatrix}$$

$$\Sigma\Sigma' = \begin{bmatrix} \Omega\Omega' & 0_{3\times1} \\ 0_{1\times3} & \sigma_\eta^2 \end{bmatrix}$$

I_4 is a four by four identity matrix and $0_{3\times1}$ and $0_{3\times1}$ are respectively a three element column and row vector of zeros. As already discussed in section 9.4, I set $\phi_y = 0.995$ (to match the value of 0.98 on annual horizon) and $\sigma_\eta^2 = 0.005$ (to match the value of 0.02 on an annual horizon). I set the intercept of the income shock $\gamma_y = -\frac{1}{2}\frac{\sigma_\eta^2}{1-\phi_y^2}$, so that the unconditional mean of $\exp(\rho_y)$ is equal to one.

The last remaining stochastic variable in the model is the idiosyncratic house price shock. Since idiosyncratic shocks are leptokurtotic (see section 3), I fit their distribution in each year and metropolitan area using a mixture of Gaussians with three components[30]. In formulas, the distribution of idiosyncratic housing shocks is modeled as:

$$\varepsilon_{fwd} \sim \sum_{k=1}^{3} p_k N\left(\mu_k, \sigma_k\right) \tag{20}$$

Tables 15, 16 and 17 in appendix A show that the mixture of Gaussians approximation closely matches centered second, third and fourth moments of house-specific shocks from the data.

In the model, the distribution of idiosyncratic shocks does not change over time. It is calibrated to match empirical estimates of house-specific shocks from the data in year t_0. This simplification relies on the notion that it would be hard for both households and econometricians to foresee changes in the amount of idiosyncratic risk or to get a sense of its future dynamics. Even from the perspective of the researcher, there is not enough history in the data to build a stochastic model for the moments of the shocks distribution.

9.8 Values of State Variables in 2001

The model is solved both for the initial year t_0 equal to 2001. The house price is calibrated to match the median house price out of all sales taking place at time t_0 in each metropolitan area.

---

[30]Mixture weights, as well as the mean and standard deviation of each component are estimated using an expectation maximization algorithm. For a detailed description of the methodology see Dempster et al. (1977).

Household income is computed in two steps. First I estimate average income for 25 to 29 years old households in year t_0 from the California CPS. I then apply the Mincer curves based on estimates from equation 18 to obtain income for households between 35 and 39 years old and between 45 and 49 years old. I further scale the income level for the city of San Francisco, to match data from the American Community Survey (ACS). The survey shows that the average income level in the San Francisco bay area is substantially higher than in the rest of California.

Table 10, shows initial real house prices (in terms of December 2000 dollars) and income level for household heads with age between 25 and 29 years, respectively for the 2001 and the 2010 calibration. Estimates are reported for each one of the three metropolitan areas analyzed in the my empirical study. The table also includes the year t_0 average values of the other exogenous state variables included in vector X. These are the annual rent-price ratio, inflation, and the real rate on 30 year fixed rate mortgages in the US. I use the rent-price ratio and the price level in each city to compute monthly rent expense. As already explained, I will take this rent value as the value of monthly housing services extracted from the house. Housing services are then kept fixed over the entire planning horizon T.

			Los Angeles				
Year	Price Level	Income Level (25-29)	RP (% annual)	$Rent$ (monthly)	inf^{1y}	$r^{30y,M}$	ρ^y
2001	$ 230,896	$ 43,378	5.11 %	$ 983	2.67 %	4.39 %	0

			San Diego				
Year	Price Level	Income Level (25-29)	RP (% annual)	$Rent$ (monthly)	inf^{1y}	$r^{30y,M}$	ρ^y
2001	$ 264,279	$ 43,378	5.48 %	$ 1,207	2.67 %	4.39 %	0

			San Francisco				
Year	Price Level	Income Level (25-29)	RP (% annual)	$Rent$ (monthly)	inf^{1y}	$r^{30y,M}$	ρ^y
2001	$ 470,502	$ 78,080	4.70 %	$ 1,843	2.67 %	4.39 %	0

Table 10: For each metropolitan area, calibration of price level, income level (both in terms of January 2000 dollars) and initial values of other exogenous states.

9.9 Bequest Coefficient and Non-Housing Wealth

I use the bequest coefficient λ to obtain a realistic of growth path for non-housing wealth over time. In order to have a comparison to the US data, I collect information from the Survey of Consumer Finances[31] (SCF) for years 1998, 2001, 2004, 2007, 2010 and 2013. Using the survey, I am able to estimate average financial (non-housing) wealth in each year and for households who

---

[31]Available for download at https://www.federalreserve.gov/econresdata/scf/scfindex.htm

are homeowners and whose household head has age from 25 to 29, 30 to 34, 35 to 39, 40 to 44, 45 to 49, 50 to 54, 55 to 59, 60 to 64, 65 to 69, 70 to 74 and greater or equal than 75. Estimates are reported in figure 19. While wealth differences across different age groups change slightly across years, there is a common pattern. US household accumulated non-housing wealth in the earlier stages of their life-cycle.
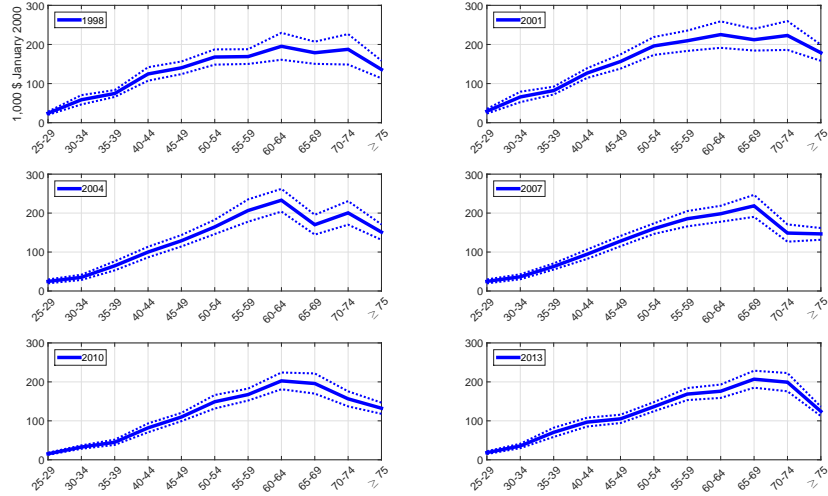

Figure 19: Average financial wealth for home-owning households by age group, based on the annual Survey of Consumer Finances for 1998, 2001, 2004, 2007, 2010 and 2013. All figures are in terms of January 2000 dollars. Standard errors take into account both imputation and sampling uncertainty are calculated following the methodology described in https://www.federalreserve.gov/Standard_Error_Documentation.pdf.

The first column of table 11 shows the increase in average non-housing wealth in between contiguous age groups. The difference between age groups is 5 years and matches the planning horizon in the model. Thus, non-housing wealth growth from the model calibrated to different age groups should match the corresponding figures from column one of table 11. I calibrate the model by setting the starting year equal to 2001 and the probability of exogenous mobility shocks π_{mob} equal to zero.

I measure wealth growth along a specific path of realizations of the exogenous state vector X. In particular, I assume that the elements of X take in each year values equal to their conditional expectation in 2001 based on the VAR in equation 19. In formulas, for each $\tau \in \{t_0 + 1, ..., t_0 + T\}$, $X_\tau = E[X_\tau | X_{t_0}]$, where expectations are based on the model in equation 19.

In the model, I set initial non-housing wealth equal to \$ 25,000 for 25-29 years old, \$ 65,000 for 35-39 years old and \$ 125,000 for 45-49 years old, matching average wealth for these groups across

the eight SCF surveys that I use in my empirical calculations[32]. The second, third and fourth column of table 11 report non-housing wealth increases from the calibrated model respectively for Los Angeles, San Diego and San Francisco. Wealth accumulation from the model matches the data closely.

	$\Delta\theta\%$ SCF	$\Delta\theta\%$ Model 2001 LA (λ)	$\Delta\theta\%$ Model 2001 SD (λ)	$\Delta\theta\%$ Model 2001 SF (λ)
$25 - 29$	93.72%	93.68% (0.70)	78.66% (0.64)	78.66% (0.64)
$30 - 34$	53.28%	-	-	-
$35 - 39$	56.80%	55.30% (0.33)	58.87% (0.37)	58.87% (0.37)
$40 - 44$	23.15%	-	-	-
$45 - 49$	27.53%	24.38 % (0.20)	22.55% (0.32)	22.55% (0.32)
$50 - 54$	14.15%	-	-	-
$55 - 59$	11.77%	-	-	-
$60 - 64$	-2.90%	-	-	-
$65 - 69$	-7.08%	-	-	-
$70 - 74$	-22.50%	-	-	-

Table 11: Increase in average non-housing wealth with respect to previous age group: Survey of Consumer Finances and model calibrated to data from Los Angeles, San Diego and San Francisco and starting year 2001. The last column shows the calibrated values of λ for households with initial age respectively between 25-29, 35-39 and 45-49. In the model, initial wealth is set respectively to $ 25,000 for 25-29, $ 65,000 for 35-39, and $ 125,000 for 45-49.

10 Quantitative Exercise

The optimization problem is solved numerically by backward induction, starting from the final period $t_0 + T$. Dynamics of the exogenous state vector X are modeled using the method developed by Gospodinov and Lkhagvasuren (2013). This methodology approximates the VAR dynamics using a Markov chain with a discrete number of states. Each grid of nodes corresponds to different realizations of the random vector, and the methodology provides values of Markov transition probabilities across nodes (for a detailed discussion, see appendix E). The results reported in this section are based on a calibration where the number of discrete Markov states is equal to 1,024. In each node, I compute continuation utilities by using conditional transition probabilities over the future dynamics of the random state vector X and (when needed) over the distribution idiosyncratic housing shocks. As already explained in section 9.7, the distribution of the idiosyncratic housing

---

[32]An issue with SCF data is that the survey overweights wealthy households in its sample, as is for instance argued at https://www.federalreserve.gov/econresdata/scf/files/isi2007.pdf. Thus, the non-housing wealth/age relationship presented in figure 19 might overestimate the increase in wealth over the lifecycle, especially for the later stages of life (households with head older than 55 years). In this sense, households propensity to save in my calibration might be overstated. Nonetheless, while a value of λ smaller than one is needed to prevent the household from consuming all of her wealth over the planning horizon, the nature of the results in the following sections does not crucially depend on the choice of λ.

shock follows a mixture of Gaussians. Stochastic integration over each component is done with quadrature, using methodology and code from Judd et al. (2011).

The model is solved separately for each metropolitan area (Los Angeles, San Diego and San Francisco) and for households belonging to different age groups (25 to 29 years old, 35 to 39 years old and 45 to 49 years old). To account for the increase in idiosyncratic risk taking place during and after the crisis, I run two different versions of the model, one calibrated so that the initial period t_0 corresponds to year 2001, and one in which t_0 is 2010. As explained in section 9.3, I run the model multiple times with different calibrated values of the annual probability of the exogenous mobility shocks. Section 9 already covered in detail the calibration of parameters, stochastic variables and other features of the model. Nonetheless, the values of most of the main parameters are once again summarized below in table 12.

		25-29 yrs	**35-39 yrs**	**45-49 yrs**
Preferences	γ	5.00	5.00	5.00
	σ	0.50	0.50	0.50
	β	0.98	0.98	0.98
	δ	0.20	0.20	0.20
Horizon	T	5 yrs	5 yrs	5 yrs
Bank Rate	r_b	2%	2%	2%
Mobility	π_{mob}	0%,12%,30%	0%,10%,25%	0%,5%,20%

Table 12: Summary of calibrated values for the main parameters in the model.

I use the model to assess the welfare cost of idiosyncratic risk. The calculation is done in two steps. First I evaluate the percentage decrease in the present value of expected utility in period t_0, due to idiosyncratic shocks. The percentage cost is then converted into the annual payment that the household would be willing to make in order to insure against house-specific risk. Welfare costs and insurance premiums are computed for households with different levels of initial non-housing wealth θ. Intuitively, homeowners who have greater financial wealth can use it to absorb the wealth shocks generated by idiosyncratic housing risk. This effect is particularly important when the household is facing exogenous mobility shocks, and is therefore forced to sell.

10.1 Welfare Cost: Baseline Model

Results from the quantitative model show that household are willing to pay high premiums to insure against idiosyncratic risk are high. In the presence of exogenous mobility shocks, premiums are larger for households with lower initial non-housing wealth. In this section I focus on the baseline framework, where there is no mortgage and the household owns the entire capital invested in the house. I first discuss results for the city of Los Angeles, for households in the three different age groups (25 to 29, 35 to 39 and the 45 to 49 years old). Results for San Diego and San Francisco are qualitatively similar and are reviewed at the end of the section.

Figure 20 reports the costs of idiosyncratic risk for the youngest age group (25 to 29 years old) in Los Angeles, when the initial time period set to 2001. Welfare costs are presented in terms of the annual insurance premium that the household would pay to remove idiosyncratic risk. The model is solved for households with different initial amounts of non-housing wealth θ. The figures show results for the range of values of θ in between 0 and \$ 300,000. For households not exposed to exogenous mobility shocks ($\pi_{mob} = 0\%$), the percentage welfare cost is flat across different values of non-housing wealth. Since value functions are increasing in initial wealth, this leads to insurance premiums that are initially increasing with θ. Households with zero financial wealth would pay approximately \$ 600 a year to insure, while households with \$ 150,000 in the bank account would pay slightly less than \$ 900. This result is driven by the fact that young households in this calibration are highly unlikely to endogenously sell their house. When exogenous mobility shocks are included, the percentage welfare cost of idiosyncratic risk becomes decreasing in non-housing wealth. For $\pi_{mob} = 12\%$, which is close to annual mobility of young homeowners in Census data, the insurance premium is slightly higher than \$ 1,000 dollars when non-housing wealth is smaller than \$ 150,000. I have shown in table 10 that monthly rent in this calibration is slightly smaller than \$ 1,000, so the insurance payment is substantial. When I set $\pi_{mob} = 30\%$, consistent with annual mobility for young renters in the data, the annual premium is \$ 1,400 dollars for household with non-housing wealth equal to zero. The premium is then monotonically decreasing in non-housing wealth, reaching a value only slightly above \$ 900 when θ is equal to \$ 300,000.

Figures 21 and 22 show welfare costs respectively for households of age in between 35 and 39 years old and in between 45 and 49 years old. The initial year is set to 2001. These households have lower expected income growth than the youngest group, but also much higher starting annual income: respectively \$ 58,750 and \$ 62,100 with respect to the \$ 43,400 of households with age in between 25 and 29. The older households also have a high propensity to accumulate wealth, according to the data used in the calibration of λ from section 9.9. These households can save enough to buy a larger house and are therefore likely to upgrade house quality.

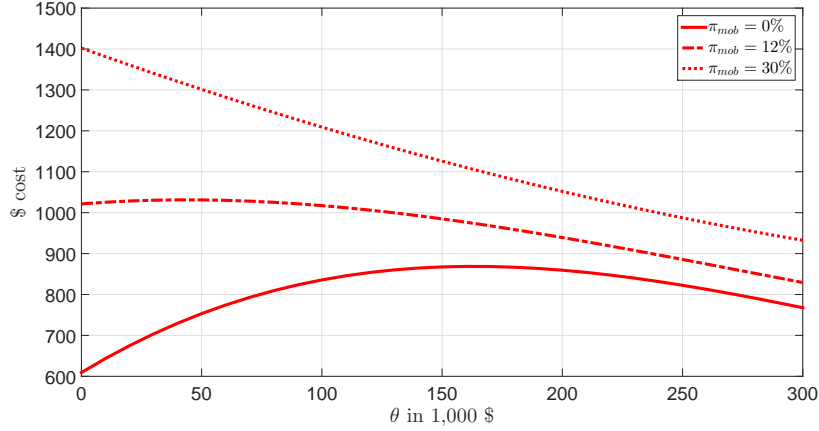Households with low initial non-housing wealth are always the most affected by idiosyncratic

Figure 20: Cost of idiosyncratic risk for 25 to 29 years old households in Los Angeles in 2001. Each line corresponds to a different level of the probability of exogenous mobility shocks π^{mob}. The right panel plots the value of the equivalent annual payment $\overline{\Delta Y}$.

risk. Figures and show that even when $\pi_{mob} = 0$, the welfare cost of idiosyncratic risk is decreasing in θ. For $\pi_{mob} = 0$, poor households are therefore willing to pay larger insurance premiums than households with same non-housing wealth, but age in between 25 and 29 years old.

Introducing mobility shocks leads to effects similar to the ones already discussed for the youngest households group. For homeowners of age in between 35 and 39 years I consider values of π_{mob} of 10% and 25%, close respectively to annual mobility from Censis for homeowners and renters in the same age group. Exogenous mobility shocks lead to higher welfare decreases and insurance premiums, especially for households with non-housing wealth below $ 100,000. For households with age between 45 and 49, I set π_{mob} equal respectively to 5% and 20%. For this last age group, once θ is fixed, the increase in insurance premiums determined by exogenous mobility shocks is smaller than for the younger groups, even when considering the poorer households. This is due to the higher income of the 45 to 49 years old households.

Insurance premiums for the metropolitan areas of San Diego and San Francisco are reported in table 13. The results in the table focus on households in the age group between 25 and 29, which, as already discussed for Los Angeles, are the most affected by idiosyncratic housing risk, due to their higher mobility and lower labor income. Since we have shown that idiosyncratic risk is particularly important for households with low non-housing wealth, the table reports insurance premiums computed when the initial value of θ is respectively set equal $ 0, $ 20,000, $ 50,000 and $ 100,000.
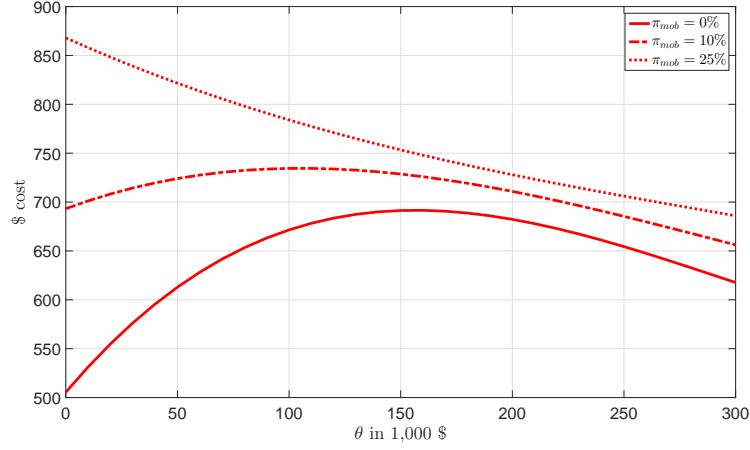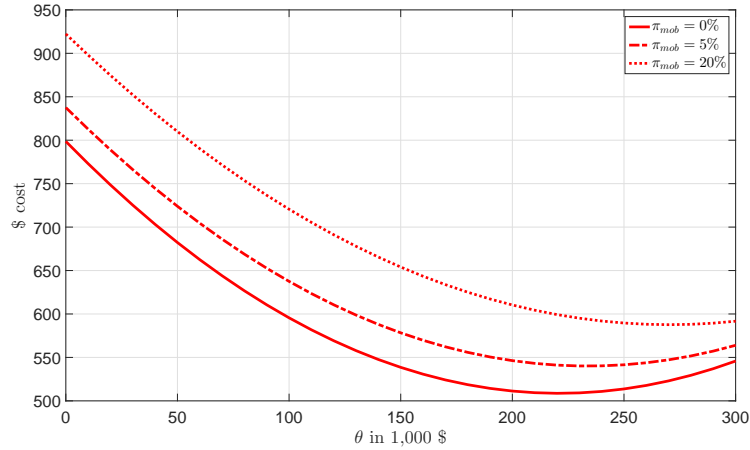
Figure 21: Cost of idiosyncratic risk for 35 to 39 years old households in Los Angeles in 2010. Each line corresponds to a different level of the probability of exogneous mobility shocks π^{mob}. The cost is expressed in terms of the equivalent annual payment $\overline{\Delta Y}$.

Figure 22: Cost of idiosyncratic risk for 45 to 49 years old households in Los Angeles in 2001. Each line corresponds to a different level of the probability of exogneous mobility shocks π^{mob}. The cost is expressed in terms of the equivalent annual payment $\overline{\Delta Y}$.

Similarly to what we have seen for Los Angeles, households with low non-housing wealth pay high insurance premiums to remove idiosyncratic risk. These premiums can exceed the value of one month of rent (that for the San Francisco calibration is slightly above 1,840 dollars. Insurance premiums are increasing in the probability of facing an exogenous mobility shock. House-specific risk in San Diego is smaller than in the other two metropolitan areas included in the study[33]. In fact, insurance premiums for this last metropolitan area are smaller than one month of rent, even when the annual probability of facing an exogenous mobility shock is equal to 30%, and the household has no money in her bank account.

| | θ | Age 25-29 | | |
		$\pi_{mob} = 0\%$	$\pi_{mob} = 12\%$	$\pi_{mob} = 30\%$
	$ 0	$ 442	$ 581	$ 749
San Diego	$ 20,000	$ 446	$ 575	$ 736
	$ 50,000	$ 450	$ 567	$ 717
	$ 100,000	$ 450	$ 551	$ 687
	θ	Age 25-29		
		$\pi_{mob} = 0\%$	$\pi_{mob} = 12\%$	$\pi_{mob} = 30\%$
	$ 0	$ 1,831	$ 2,102	$ 2,364
San Francisco	$ 20,000	$ 1,570	$ 1,820	$ 2,065
	$ 50,000	$ 1,274	$ 1,497	$ 1,720
	$ 100,000	$ 1,008	$ 1,194	$ 1,387

Table 13: Insurance premiums for the cities of San Diego and San Francisco. Results are based on a calibration that sets t_0 equal to 2001, and refer to households in the age group in between 25 and 29 years old.

10.2 Debt and Default

I now introduce debt into the model, modeling the mortgage as a 30 year nominal FRM. Maturity in the model matters for the magnitude of mortgage installments, that I defined as I_M in section 8.4. For the 30 year mortgage, the installment will be equal to 1/30 of the amount initially borrowed. As explained in section 9.6 the nominal mortgage rate is fixed at the time debt is issued, but the real rate is then exposed to fluctuations in inflation.

At time t_0 (2001) the household has a mortgage with initial loan-to-value ratio equal to 80%. In other words, initially the household directly owns only 20% of the capital invested in the house. Each period she makes a payment consisting of an installment of the mortgage notional and interest on the remaining mortgage amount. When selling her home, the household pays back the remaining mortgage and is allowed to lever up again, with maximum loan-to-value ratio (η in the model from section 8.4) equal to 90%.

---

[33]See for example evidence reported in figure 16 from section 7, and figure 3 in section 3.

Figure 23, in the same fashion as figure 20, reports insurance premiums for the metropolitan area of Los Angeles and households in the youngest age group (25 to 29 years old). Premiums are plotted across different levels of initial non-housing wealth θ and for different levels of the probability of facing exogenous mobility shocks π_{mob}. With respect to the model without leverage, insurance premiums are now substantially larger for homeowners with non-housing wealth smaller than $ 100,000. Even when π_{mob} is set equal to zero, households with non-housing wealth smaller than $ 50,000 are willing to pay insurance premiums larger than one month of rent. When initial non-housing wealth is equal to zero and pi_{mob} is set to 12%, the annual insurance premium is equal to approximately $ 1,550 (in the Los Angeles calibration, one month of rent/housing services is equal to approximately $ 983). In general, the fact that the household now only owns 20% of the capital invested in the house makes the impact of idiosyncratic shocks more relevant.
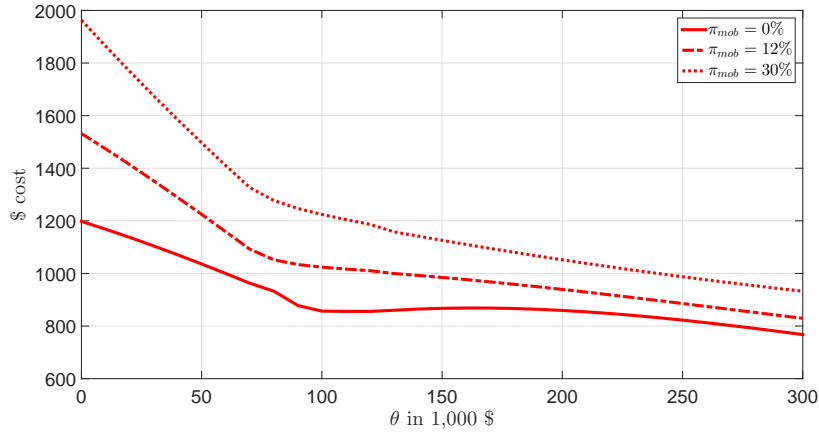

Figure 23: Cost of idiosyncratic risk for 25 to 29 years old households in Los Angeles in 2001. Each line corresponds to a different level of the probability of exogeneous mobility shocks π^{mob}. The cost is expressed in terms of the equivalent annual payment $\overline{\Delta Y}$.

I have shown, using the Los Angeles calibration, that insurance premiums increase when leverage is introduced into the model. The effect is particularly evident for households with low non-housing wealth. Qualitatively similar results hold for the other two metropolitan areas included in this study. Table 14 reports values of insurance premiums for San Diego and San Francisco, and for households with age in between 25 and 29 years old. In San Diego, when π_{mob} is set equal to 12% and θ equal to $ 0, the premium is larger than two thirds of a month of rent. In San Francisco, when θ is equal to $ 0 or $ 20,000, insurance premiums are larger than one month of rent even when π_{mob} is equal to zero.

	θ	Age 25-29		
		$\pi_{mob} = 0\%$	$\pi_{mob} = 12\%$	$\pi_{mob} = 30\%$
San Diego	$ 0	$ 593	$ 819	$ 1,007
	$ 20,000	$ 559	$ 752	$ 926
	$ 50,000	$ 507	$ 654	$ 809
	$ 100,000	$ 453	$ 554	$ 693
	θ	Age 25-29		
		$\pi_{mob} = 0\%$	$\pi_{mob} = 12\%$	$\pi_{mob} = 30\%$
San Francisco	$ 0	$ 2,603	$ 3,153	$ 3,309
	$ 20,000	$ 2,107	$ 2,502	$ 2,684
	$ 50,000	$ 1,554	$ 1,778	$ 1,977
	$ 100,000	$ 1,066	$ 1,227	$ 1,423

Table 14: Insurance premiums from the version of the model allowing for debt financing, and for the cities of San Diego and San Francisco. Results are based on a calibration that sets t_0 equal to 2001, and refer to households in the age group in between 25 and 29 years old.

11 Concluding Remarks and Future Work

The analysis in this paper provides three main contributions. The first one consists of the data and methodology used to measure idiosyncratic housing risk. I rely on a novel dataset, which includes remodeling expenses, as well as methodology from the private equity literature to estimate house-specific risk in the metropolitan areas of Los Angeles, San Diego and San Francisco. The second contribution, and key empirical result, is the finding that idiosyncratic risk does not increase with holding period. An implication of this result is that the relative importance of the house-specific component of house price risk decreases as a home holding period becomes longer. The last contribution is the quantitative assessment of the welfare cost of idiosyncratic house risk. I build a quantitative portfolio model where I take the perspective of a homeowner who may sell her house due to an endogenously optimal decision or an exogenously determined mobility shock. I find that homeowners would be willing to pay sizable premiums in order to insure against idiosyncratic risk. Premiums are substantially higher for households that are more likely to face exogenous mobility. Idiosyncratic housing risk also has its greatest impact on homeowners that have little non-housing wealth. The largest premiums are then paid by young households who have lower income, are less wealthy and more likely to be affected by exogenous mobility shocks.

Several lines of future work can spawn from this project. First, it is relevant to explore how the young households choice between owning and renting is affected by idiosyncratic risk. In fact, due to idiosyncratic risk, young families may find it optimal to postpone ownership to a time when they have built up more wealth and are less likely to face mobility shocks.

A second direction of future work is linking the magnitude of idiosyncratic risk and its term structure properties to the micro-structure of housing markets. The flat term structure of house-

specific risk may be determined by quality shocks with particular properties. However, an alternative explanation is that idiosyncratic risk is driven by price risk at the time of sale. We know from previous literature that housing markets are illiquid and segmented. Search behavior, information asymmetries and bargaining are key determinants of price formation, as shown by Piazzesi and Schneider (2009b), Genesove and Han (2012), Piazzesi et al. (2014), Kurlat and Stroebel (2015) and several other studies reviewed by Han and Strange (2014). Thus, a large fraction of idiosyncratic risk may be driven by heterogeneity in bargaining power and search costs across market participants at the time of sale. Further empirical and modeling work could be pursued by merging the dataset used in this paper with information on home price listings. This information should include changes in listed prices after the first listing, as well as the time spent by a property on the market before sale. The empirical study would give some insights into the bargaining power of sellers and buyers. The housing literature has already used listings data to explore market segmentation within metropolitan areas, as is done for example by Piazzesi et al. (2014), who work with data from the metropolitan area of San Francisco.

In the absence of listings data, further work could be based on methodology from industrial organization research. There is a literature that has developed methods to extract information on heterogeneous frictions faced by buyers using price dispersion in markets for homogeneous non-durable goods. Hong and Shum (2006) focus on an environment with one sided search and, given assumptions on the form of search and market characteristics, can infer the distribution of search costs. Similar frameworks can be found in Moraga-Gonzalez and Wildenbeest (2008) and Moraga-Gonzalez et al. (2012). Note that since this literature is concerned with non-durable goods markets, buyers are the only side facing search costs. Sellers only bear the marginal cost of providing the good, which is assumed to be either homogeneous across sellers or exogenously determined. Thus, an issue with implementing these frameworks in housing is that there is search both on the home seller and the home buyer side. Since heterogeneous frictions affect both counter-parties, it is likely that the distribution of prices would not be enough to infer properties of search costs.

Finally, an interesting topic for future research are the benefits of more efficient and liquid housing markets. Under the assumption that idiosyncratic risk is mostly driven by housing market micro-structure, the insurance premiums generated by my model measure the cost of market inefficiency from the perspective of homeowners. A regulator might be willing to devote resources to improve market efficiency, for example by supporting centralized trading and dealers acting as intermediaries in residential real estate markets.

References

Alntonji, Joseph G. and Lewis M. Segal, "Small Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business Economics and Statistics*, 1996.

Anenberg, Elliot and Edward Kung, "Estimates of the Size and Source of the Price Declines Due to Nearby Foreclosures," *American Economic Review*, 2014.

Attanasio, Orazio, Andrew Leicester, and Matthew Wakefield, "Do House Prices Drive Consumption Growth? The Coincident Cycles of House Prices and Consumption in the UK," *Journal of the European Economic Association*, 2011, *9* (3), 399–435.

Attanasio, Orazio P. and Guglielmo Weber, "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey," *Journal of Political Economy*, 1995.

Axelson, Ulf, Morten Sorensen, and Per Stromberg, "Alpha and Beta of Buyout Deals: A Jump CAPM for Long-Term Illiquid Investments," *Working Paper*, 2015.

Badarinza, Cristian, John Y. Campbell, and Tarun Ramadorai, "International Comparative Household Finance," *Working Paper*, 2016.

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

Campbell, John Y. and Joao F. Cocco, "Household Risk Management and Optimal Portfolio Choice," *Quantitative Journal of Economics*, 2003.

Campbell, John Y and Joao F. Cocco, "How Do House Prices Affect Consumption? Evidence from Micro Data," *Journal of Monetary Economics*, 2007.

Campbell, John Y. and Joao F. Cocco, "A Model of Mortgage Default," Working Paper 17516, National Bureau of Economic Research 2011.

_ , Martin Lettau, Burton G. Malkiel, and Yexiao Xu, "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *Journal of Finance*, 2001.

Cao, Charles, Timothy Simin, and Jing Zhao, "Can Growth Options Explain the Trend in Idiosyncratic Risk?," *Review of Financial Studies*, 2006.

Case, Karl E. and Robert J. Shiller, "Price of Single Family Homes Since 1970: New Indexes for Four Cities," Working Paper 2393, National Bureau of Economic Research 1987.

_ **and** _ , "The Efficiency of the Market for Single Family Homes," Working Paper 2506, National Bureau of Economic Research 1988.

Chen, Hui, Michael Michaux, and Nikolai L. Roussanov, "Houses as ATMs? Mortgage Refinancing and Macroeconomic Uncertainty," SSRN Scholarly Paper ID 2024392, Social Science Research Network, Rochester, NY 2013.

Chetty, Raj, Adam Szeidl, and Laszlo Sandor, "The Effect of Housing on Portfolio Choice," Working Paper 15998, National Bureau of Economic Research 2016.

Cocco, Joao F., "Consumption and Portfolio Choice over the Life Cycle," *Review of Financial Studies*, 2005.

_ , "Portfolio Choice in the Presence of Housing," *Review of Financial Studies*, 2005, *18* (2), 535–567.

Davis, Morris A. and Michael G. Palumbo, "The Price of Residential Land in Large US Cities," *Journal of Urban Economics*, 2006.

Dempster, A. P., N. M. Laird, and Rubin D. B., "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, 1977.

Duffie, Darrel, Nicolae Garleanu, and Lasse H. Pedersen, "Over-the-Counter-Markets," *Econometrica*, 2005.

Englund, Peter, John M. Quigley, and Christian L. Redfearn, "Improved Price Indexes for Real Estate: Measuring the Course of Swedish Housing Prices," *Journal of Urban Economics*, 1998.

Epstein, Larry, Fahri Emmanuel, and Gita Gopinath, "How Much Would You Pay to Resolve Long-Run Risk?," *American Economic Review*, 2014.

Epstein, Larry G. and Stanley E. Zin, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 1989, *57* (4), 937–969.

Fama, Eugene F. and Kenneth R. French, "A Five-Factor Asset Pricing Model," *Journal of Financial Economics*, 2014.

Favilukis, Jack, Sydney C. Ludvigson, and Stijn Van Nieuwerburgh, "The Macroeconomic Effects of Housing Wealth, Housing Finance, and Limited Risk-Sharing in General Equilibrium," Working Paper 15988, National Bureau of Economic Research 2010.

Ferreira, Fernando Vendramel, "You Can Take It with You: Transferability of Proposition 13 Tax Benefits, Residential Mobility and Willingness to Pay for Housing Amenities," Working Paper 72, Center for Labor Economics UC Berkeley 2004.

Flavin, Marjorie and Takashi Yamashita, "Owner-Occupied Housing and the Composition of the Household Portfolio," *American Economic Review*, 2002, *92* (1), 345–362.

Gan, Jie, "Housing Wealth and Consumption Growth: Evidence from a Large Panel of Households," *Review of Financial Studies*, 2010.

Genesove, David and Lu Han, "Search and Matching in the Housing Market," *Journal of Urban Economics*, 2012.

Giacoletti, Marco and Victor Westrupp, "Residential Real Estate Traders: Risk, Returns and Strategies," *Working Paper*, 2016.

Goetzmann, William N., "The Single Family Home in the Investment Portfolio," *The Journal of Real Estate Finance and Economics*, 1993.

_ **and Matthew Spiegel**, "Non-Temporal Components of Residential Real Estate Appreciation," *The Review of Economics and Statistics*, 1995.

Gospodinov, Nikolay and Damba Lkhagvasuren, "A Moment Matching Method for Approximating Vector Autoregressive Processes by Finite State Markov Chains," *Journal of Applied Econometrics*, 2013.

Hall, Robert E., "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 1988.

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.

Hansen, Lars Peter and Kenneth J. Singleton, "Generalize Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 1982.

_ **and** _ , "Efficient Estimation of Linear Asset-Pricing Models with Moving Average Errors," *Journal of Business Economics & Statistics*, 1996.

Harding, John P., Stuart S. Rosenthal, and C. F. Sirmans, "Depreciation of Housing Capital, Maintenance and House Price Inflation: Estimates from a Repeated Sales Model," *Journal of Urban Economics*, 2006.

Hong, Han and Matthew Shum, "Using Price Distributions to Estimate Search Costs," *RAND Journal of Economics*, 2006.

Iacoviello, Matteo, "Housing Wealth and Consumption," Working Paper, Federal Reserve Board 2011.

Ihrke, David K. and Carol S. Faber, "Geography Mobility: 2005 to 2010," Working Paper, US Department of Commerce 2012.

Judd, Kenneth L., Lilia Maliar, and Serguei Maliar, "Numerically Stable and Accurate Stochastic Simulation Approaches for Solving Dynamic Economic Models," *Quantitative Economics*, 2011.

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

Knight, John R. and C. F. Sirmans, "Depreciation, Maintenance and House Prices," *Journal of Housing Economics*, 1996.

Korteweg, Arthur and Stegan Nagel, "Risk Adjusting the Returns to Venture Capital," *Journal of Finance*, 2015.

Kurlat, Pablo and Johannes Stroebel, "Testing for Information Asymmetries in Real Estate Markets," *Review of Financial Studies*, 2015.

Lu, Han, "The Effects of Price Risk on Housing Demand: Empirical Evidence from US Markets," *Review of Financial Studies*, 2010.

_ , "Understanding the Puzzling Risk-Return Relationship for Housing," *Review of Financial Studies*, 2013.

Meghir, Costas and Luigi Pistaferri, "Income Variance Dynamics and Heterogeneity," *Econometrica*, 2004.

_ **and** _ , "Earnings, Consumption and Lifecycle Choices," *Handbook of Labor Economics*, 2011.

Moraga-Gonzalez, Jose L. and Matthijs R. Wildenbeest, "Maximum Likelihood Estimation of Search Costs," *European Economic Review*, 2008.

_ **, Sandor Zsolt, and Matthijs R. Wildenbeest**, "Semi-Non Parametric Estimation of Consumer Search Costs," *Journal of Applied Econometrics*, 2012.

Moran, P. A. P., "Notes on Continuous Stochastic Phenomena," *Biometrika*, 1950.

Peng, Lian, "The Risk and Return of Commercial Real Estate: A Property Level Analysis," *Real Estate Economics*, 2015.

Phillips, Peter C. B. and Ke-Li Xu, "Tilted Nonparametric Estimation of Volatility Functions with Empirical Applications," Discussion Paper 1612, Cowles Foundation 2011.

Piazzesi, Monika and Martin Schneider, "Inflation and the Price of Real Assets," SSRN Scholarly Paper ID 1425164, Social Science Research Network, Rochester, NY 2009.

_ **and** _ , "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 Selale Tuzel**, "Housing Consumption and Asset Pricing," *Journal of Financial Economics*, 2007.

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

Rouwenhorst, Geert K., *Asset Pricing Implications of Wage and Employment Dynamics*, Princeton University Press, 1995.

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

Saporta-Eksten, Itay, "Job Loss, Consumption and Unemployment Insurance," *Working Paper*, 2014.

Silverman, B., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, 1984.

Sinai, Todd and Nicholas S. Souleles, "Owner-Occupied Housing as a Hedge Against Rent Risk," *Quantitative Journal of Economics*, 2005.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron, "Consumption and Risk Sharing over the Life Cycle," *Journal of Monetary Economics*, 2004.

Vissing-Jorgensen, Annette and Orazio P. Attanasio, "Stock-Market Participation, Intertemporal Substitution, and Risk Aversion," *American Economic Review*, 2003.

Wilhelmsson, Mats, "House price depreciation rates and level of maintenance," *Journal of Housing Economics*, 2007.

Yao, Rui and Harold H. Zhang, "Optimal Consumption and Portfolio Choice with Risky Housing and Borrowing Constraints," *Review of Financial Studies*, 2005.

A   Additional Tables

	Data			Simulation		
	Los Angeles	San Diego	San Francisco	Los Angeles	San Diego	San Francisco
2001	0.12	0.08	0.11	0.12	0.08	0.11
2002	0.12	0.08	0.10	0.12	0.08	0.10
2003	0.12	0.08	0.10	0.12	0.08	0.10
2004	0.11	0.09	0.11	0.11	0.09	0.11
2005	0.11	0.08	0.10	0.11	0.08	0.10
2006	0.10	0.09	0.10	0.10	0.09	0.10
2007	0.12	0.09	0.11	0.12	0.09	0.11
2008	0.13	0.10	0.13	0.13	0.10	0.13
2009	0.13	0.09	0.14	0.13	0.09	0.14
2010	0.12	0.10	0.13	0.12	0.10	0.13
2011	0.13	0.10	0.14	0.13	0.10	0.14
2012	0.13	0.10	0.14	0.13	0.10	0.14

Table 15: Standard deviation of residuals by year and city. The first three columns refer to the estimated residuals from regression equation 2, the columns from the fourth to the sixth refer to residuals simulated from the mixture of Gaussians in 20.

	Data			Simulation		
	Los Angeles	San Diego	San Francisco	Los Angeles	San Diego	San Francisco
2001	-0.12	0.25	-0.75	-0.12	0.24	-0.78
2002	-0.22	0.02	-0.51	-0.17	0.10	-0.53
2003	-0.38	0.13	-0.70	-0.38	-0.21	-0.44
2004	-0.45	-0.42	-0.24	-0.47	-0.32	-0.10
2005	-0.36	-0.03	-0.45	-0.25	0.05	-0.31
2006	0.14	0.29	0.03	0.09	0.26	0.08
2007	0.22	0.24	-0.04	0.25	0.24	-0.02
2008	-0.02	-0.08	-0.22	-0.04	-0.09	-0.24
2009	-0.13	-0.22	-0.18	-0.11	-0.22	-0.17
2010	-0.16	-0.33	-0.28	-0.17	-0.32	-0.29
2011	-0.03	-0.02	-0.09	-0.05	-0.00	-0.12
2012	-0.39	-0.20	-0.32	-0.37	-0.21	-0.29

Table 16: Skewness of residuals by year and city. The first three columns refer to the estimated residuals from regression equation 2, the columns from the fourth to the sixth refer to residuals simulated from the mixture of Gaussians in 20.

|	Data			Simulation		
	Los Angeles	San Diego	San Francisco	Los Angeles	San Diego	San Francisco
2001	5.38	4.69	7.81	5.32	4.54	8.39
2002	5.90	5.18	8.82	5.87	4.59	9.30
2003	5.52	10.76	10.50	5.69	8.78	10.34
2004	5.98	5.86	7.10	5.98	5.02	7.05
2005	6.92	5.10	7.55	6.59	4.74	7.18
2006	5.62	4.21	6.19	5.44	4.02	6.10
2007	4.68	3.51	5.44	4.58	3.52	5.38
2008	4.21	3.28	4.42	4.28	3.33	4.57
2009	4.50	3.27	4.86	4.34	3.24	5.04
2010	4.43	4.20	5.30	4.34	4.06	5.16
2011	4.09	3.42	4.65	4.09	3.32	4.55
2012	4.19	3.58	4.91	4.19	3.57	4.74

Table 17: Kurtosis of residuals by year and city. The first three columns refer to the estimated residuals from regression equation 2, the columns from the fourth to the sixth refer to residuals simulated from the mixture of Gaussians in 20.

B Additional Figures
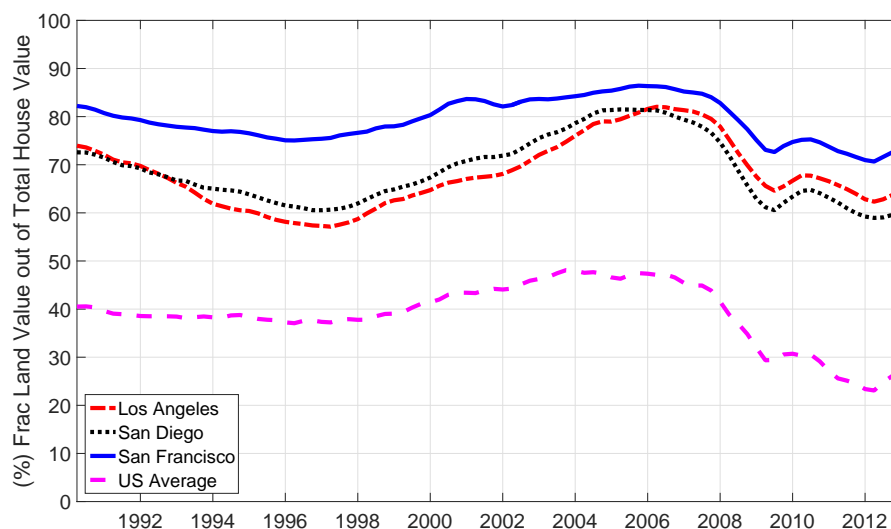
Figure B.24: Fraction of land value out of total house value. I report time series for the metropolitan areas of Los Angeles, San Diego, San Francisco (which is computed as the average between the fraction for the metropolitan area of San Francisco and San Jose), as well as for the average fraction across all US metropolitan areas. The sample is quarterly, from 1990 Q1 to 2012 Q4.
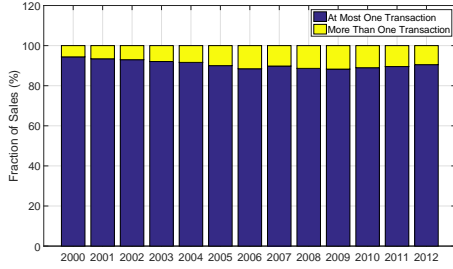
(a) Number of trades for 2 years before transaction
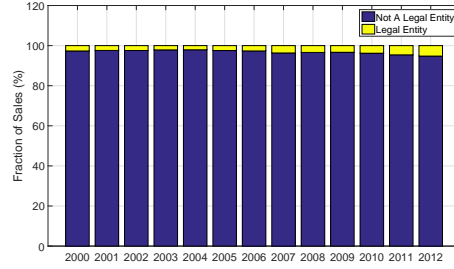
(b) Is seller a legal entity?

Figure B.25: Fraction of resales carried out by agents likely to be professional investors in each year between 2001 and 2012. Panel (a) shows the fraction of sales by agents that were involved in at least two other house transactions in the two previous years. Panel (b) shows the fraction of sales by agent who are legal entities.
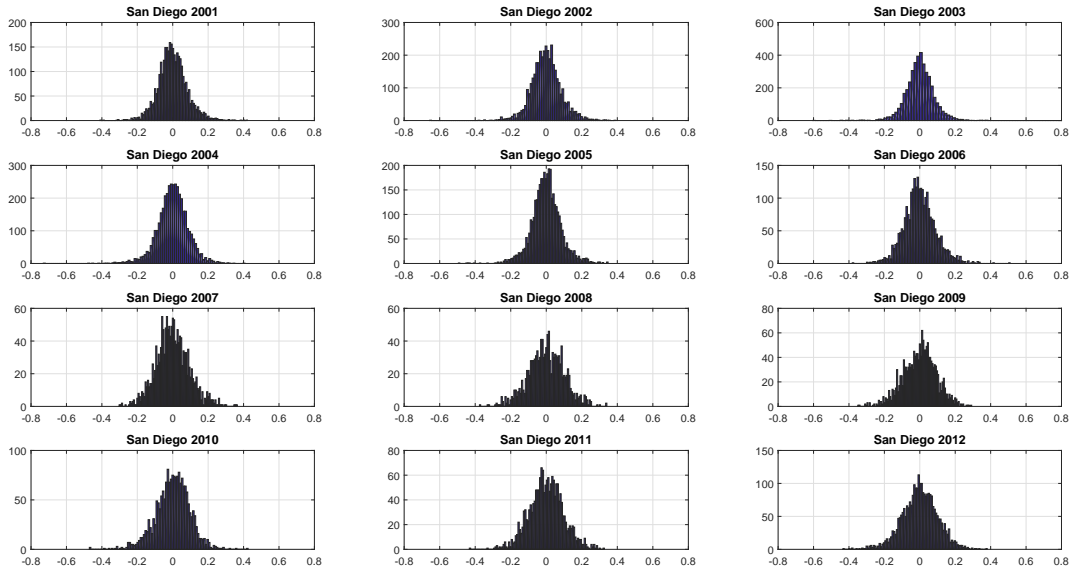

Figure B.26: Distribution of residuals from equation 2, years from 2001 to 2012, San Diego.
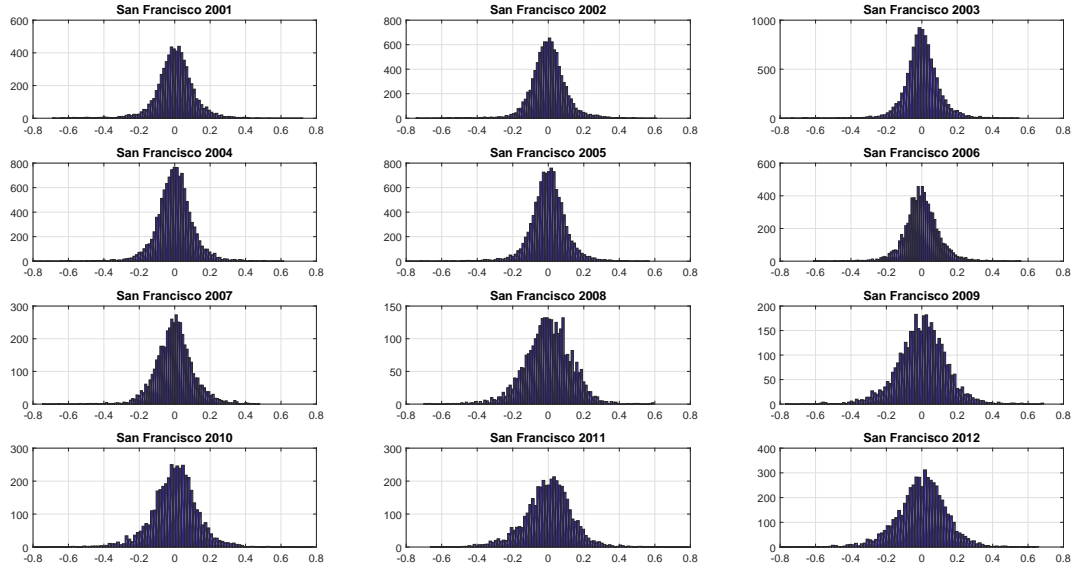
Figure B.27: Distribution of residuals from equation 2, years from 2001 to 2012, San Francisco.
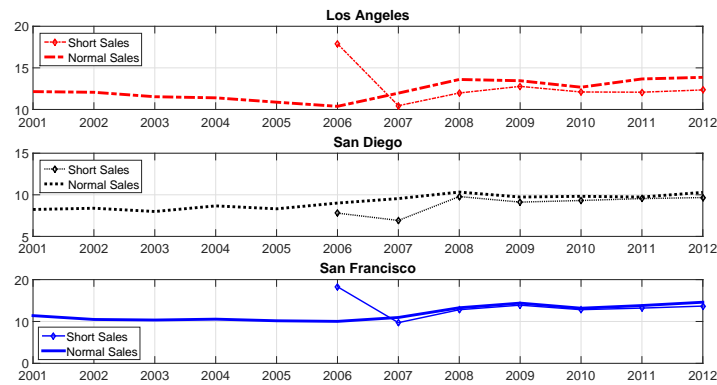
Figure B.28: Standard deviation of residuals from equation 2, respectively for short sales and normal sales, city by city; years from 2001 to 2012.
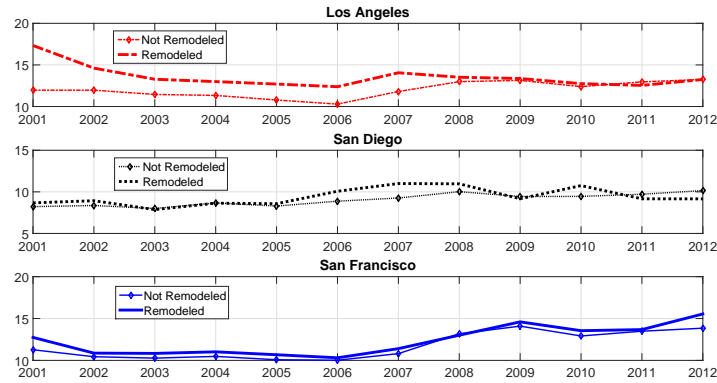
Figure B.29: Standard deviation of residuals from equation 2, respectively for houses that were and were not remodeled, city by city; years from 2001 to 2012.
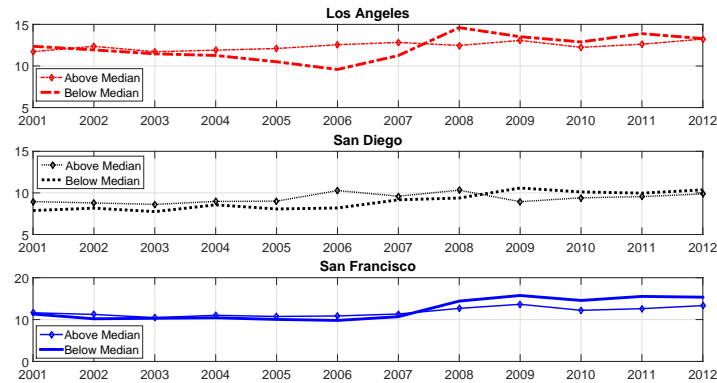

Figure B.30: Standard deviation of residuals from equation 2, respectively for houses that were bought at a price above and below the median house price in their city in the year they were bought, city by city; years from 2001 to 2012.

C   Data Methodology

C.1   CoreLogic Deed Data

In this subsection, I 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, I 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*[34]. Moreover, I 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, I treat (and remove) duplicated observations. We developed a variable that flags potential duplicated observations in sample. Below I 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))

---

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

- "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)

- "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, I keep only one observation per record. For all the other cases, I 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), I keep such observations in order to properly identify whether a return was generated over arms-length transactions or not. After comparing repeated sales, I remove any returns involving transfers.

Lastly, I 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.

C.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.

C.3   Merging Corelogic and Buildzoom

At this stage, I match the data between CoreLogic and Buildzoom by comparing address location under the same city. Due to potential misreporting/misspelling of city names, I 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[35]. When comparing properties

---

[35]Shapefile source: nhgis.

that were remodeled against those that were not, I keep only cities in which I 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, I 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[36].

Secondly, I 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.

D Sales Density Heat Maps

Figures D.31, D.32 and D.33 show heat maps based on sales density around each house transaction, respectively in Los Angeles, San Diego and San Francisco and for years 2001, 2004, 2008 and 2012.

Density is evaluated using a two-dimensional kernel estimator estimated over the latitude and longitude of trades. First, I normalize longitude and latitude in each metropolitan area and year so that they have variance equal to one. Then, I collect longitude and latitude of sale i taking place in year t_i into the row vector X_i and evaluate density as:

$$\hat{f}_h(X_i) = \frac{1}{N_{t_i} h^d} \sum_{j=1}^{N_{T_i}} K_h\left(X_j - X_i\right)$$

Where $d = 2$ and N_{T_i} is the total number of sales that took place in year T_i in the metropolitan if i. I set $K_h(Z) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2} Z' H Z\right)$, where H is a 2×2 matrix with h^2 on the main diagonal and zeros in the off diagonal elements. The bandwidth is chosen to match the optimality criterion for normal kernels in Silverman (1984) and is then equal to $\left(\frac{4}{5}\right)^{1/6} N_{T_i}^{-1/6}$.

In density estimation I use all sales, including REOs, since Anenberg and Kung (2014) document that while these houses sell at a discount due to worst maintenance and quality, their "competitive" effect on nearby houses on sale is similar to the one of equivalent non-REO sales. I then overlay on the heat maps dots corresponding to the locations of resales for each year.

---

[36]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.

We can clearly see in the figures below that resales are widely spread across metropolitan areas and are present in both high and low sale density neighborhoods.

Similar findings hold for years 2002, 2003, 2005, 2006, 2007, 2009, 2010 and 2011. Maps for these years are available upon request.
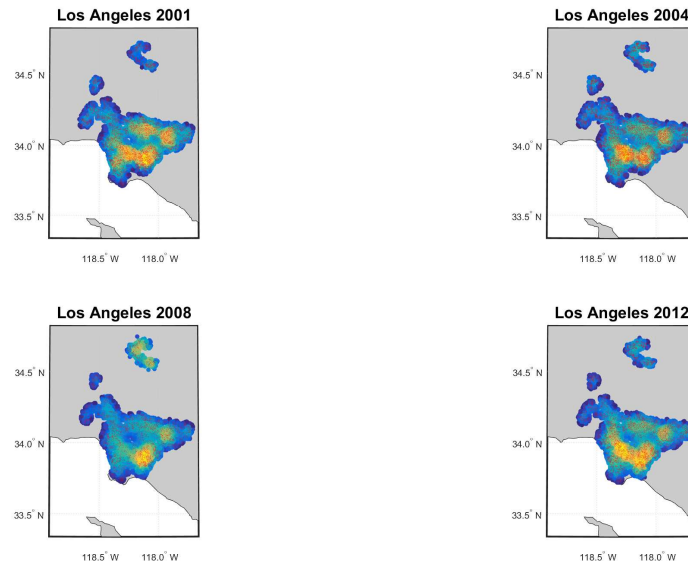

Figure D.31: Heat maps of sales frequency in Los Angeles respectively for years 2001, 2004, 2008 and 2012. Heat maps are based on all sales, while the dots represent the locations of resales used in my study.
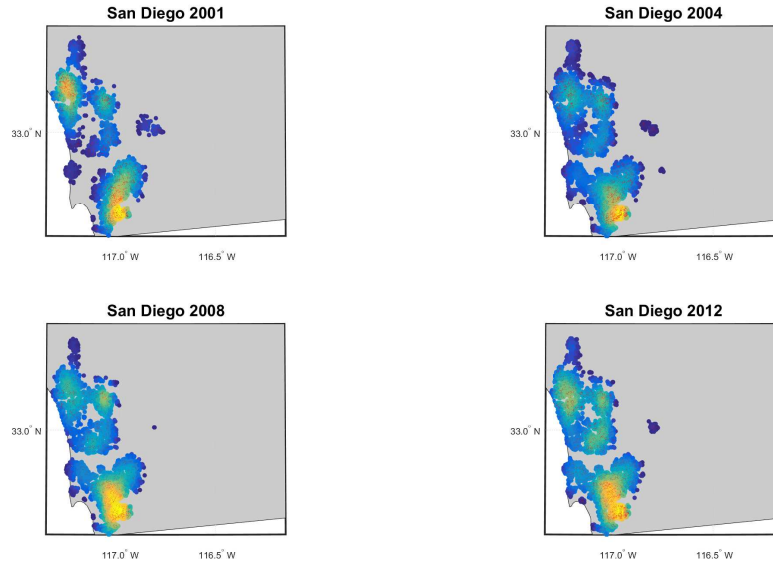
Figure D.32: Heat maps of sales frequency in San Diego respectively for years 2001, 2004, 2008 and 2012. Heat maps are based on all sales, while the dots represent the locations of resales used in my study.
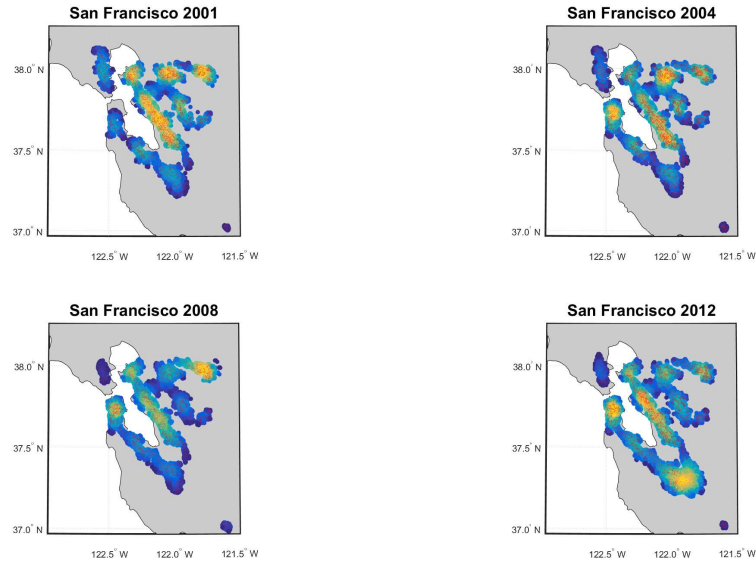

Figure D.33: Heat maps of sales frequency in San Francisco respectively for years 2001, 2004, 2008 and 2012. Heat maps are based on all sales, while the dots represent the locations of resales used in my study.

E Moment-Matching Method for Approximating VARs

The state vector follows the dynamics:

$$X_{t+1} = \Phi_0 + \Phi_1 X_t + \Sigma e_{t+1}$$

I also assume that $e_{t+1} \sim N(0, I)$. I rotate X_{t+1} into Y_{t+1} such that:

$$Y_{t+1} = A Y_t + e_{t+1}$$

Where $A = \Sigma^{-1} \Phi_1 \Sigma$ and $Y_t = \Sigma^{-1} \left(X_t - (I - \Phi_1)^{-1} \Phi_0 \right)$. Rouwenhorst (1995) develops a method to generate a discrete state space approximation of an an AR(1) process like:

$$z_{t+1} = \rho_z z_t + \sigma_{\epsilon,z} \epsilon_{z,t+1}$$

In particular, given a desired number of states N the methodology allows to calculate the desired discrete states $z^{(n)}(N, \sigma_z)$ (where σ_z is the unconditional standard deviation of z) with $n \in \{1, ..., N\}$ and transition probabilities vectors $\pi_n(N, \rho_z)$ for a discrete process $\tilde{z}$ that has the mean and the same variance as z.

Gospodinov and Lkhagvasuren (2013) generates multiple discrete distributions for each element in the stochastic vector Y_{t+1} using the methodology in Rouwenhorst (1995) and mixes them to match moments of the more general VAR process.

For $y_{k,t+1}$, kth element of Y_{t+1}, I set the number of states N_k and set $\rho_k = \sqrt{1 - 1/\sigma_{y,k}^2}$, with $\sigma_{y,k}^2 = Var(y_{k,t+1})$. The discrete states are obtained as $y_k^{(n)}(N_k, \sigma_{y,k})$ for $n \in \{1, ..., N_k\}$ as in Rouwenhorst (1995). This is gives us a state space with total number of states equal to $N^* = N_1 \times N_2 \times ... \times N_K$, where K is the number of elements in Y_{t+1}. In state $j \in \{1, ..., N^*\}$, I have a vector of discrete realizations of the elements of Y, $Y^j = \left[y_1^{(n(j,1))}, ..., y_K^{(n(j,k))} \right]$. The unknown transition probabilities from state j to state $n(l, k)$ for each element k are:

$$h_k(n(l,k), j) = Pr\left(y_{k,t} = y_k^{(n(l,k))} | Y_{t-1} = Y^{(l)} \right)$$

Where $l \in \{1, ..., N^*\}$ and $n(l, k) \in \{1, ..., N_k\}$. We can collect the probabilities in the vector

$h_k(j) = [h_k(1, j), ..., h_k(N_k, j)]$. The probability of the entire state vector moving from j to l is:

$$h(l, j) = \prod_{k=1}^{K} h_k(n(l, k), j)$$

We now want to construct the unknown transition matrix using the moment matching method. As a first step, I know that the expected value of y_k conditional on being in state j is:

$$\mu_k(j) = a_{k,1} y_1^{(n(j,1))} + ... + a_{k,K} y_K^{(n(j,K))}$$

Two different cases need to be considered:

1. If $\mu_k(j) \in \left[\rho_k y_k^{(1)}, \rho_k y_k^{(N_k)}\right]$, I start by defining $r_{i,j} = \rho_i$. Following Gospodinov and Lkhagva-suren (2013), I consider the mixture distribution:

$$\tilde{\pi}(N_i, r_{i,j}) = \lambda(r_{i,j}) \pi_n(N_i, r_{i,j}) + (1 - \lambda(r_{i,j})) \pi_{n+1}(N_i, r_{i,j})$$

Where $\pi_n(N_i, r_{i,j})$ are transition probabilities calculated according to Rouwenhorst (1995) and n is such that $r_{i,j} y_k^{(n)} \leq \mu_k(j) \leq r_{i,j} y_k^{(n+1)}$. I set:

$$\lambda(r_{i,j}) = \frac{r_{i,j} y_k^{(n+1)} - \mu_k(j)}{r_{i,j} y_k^{(n+1)} - r_{i,j} y_k^{(n)}}$$

The mixture distribution has mean $\mu_k(j)$ and variance:

$$\omega^2(r_{i,j}) = \sigma_{y,k}^2 \left(1 - r_{i,j}^2\right) + \sigma_{y,k}^2 r_{i,j}^2 \frac{4\lambda(r_{i,j})(1 - \lambda(r_{i,j}))}{N_i - 1}$$

If I find $0 < \lambda(r_{i,j}) < 1$ I solve:

$$r_{i,j}^* = \arg\min_{r_{i,j}} |\omega^2(r_{i,j}) - 1|$$

$$s.t.$$

$$r_{i,j} \in (\rho_k, 1)$$

Where I pick $r_{i,j}^*$ that satisfies the constraint and minimizes the distance between $\omega^2(r_{i,j})$ and 1, which is the conditional variance of $y_{k,t+1}$ given Y_t. I then set transition probability equal to $h_k(j) = \tilde{\pi}_n(N_k, r_{i,j}^*)$.

In the simpler cases when $\mu_i(j) = r_{i,j} y_k^{(n)}$ or $\mu_i(j) = r_{i,j} y_k^{(n+1)}$ I set either $h_k(j) = \pi_n(N_k, r_{i,j})$ or $h_k(j) = \pi_{n+1}(N_k, r_{i,j})$.

2. If $\mu_k(j) \notin \left[\rho_k y_k^{(1)}, \rho_k y_k^{(N_k)} \right]$ I set $h_k(j) = \pi_1(N_k, \rho_k)$ when $\mu_i(j) < \rho_k y_k^{(1)}$ and $h_k(j) = \pi_{N_k}(N_k, \rho_k)$ when $\mu_i(j) > \rho_k y_k^{(N_k)}$. In both cases variance is matched exactly, while the conditional mean is as close as possible given the grid points.