# Collateral constraints, wealth effects, and volatility: Evidence from real estate markets *

Barney Hartman-Glaser
UCLA

William Mann
UCLA

December 12, 2017

**Abstract**

We find that housing return volatility is negatively correlated with income at the zip-code level. We rationalize this finding with a model featuring a collateral constraint that translates income volatility to housing return volatility. Collateral constraints are tighter for lower-income areas, causing higher housing return volatility. We validate this mechanism using variation in wealth induced by lagged housing returns, using cross-sectional data on the housing expenditure share, and using state-level non-recourse status to instrument for collateral constraints. Consistent with our model, housing return volatility is negatively correlated with lagged returns, positively correlated with expenditure share, and higher in non-recourse states.

*barney.hartman-glaser@anderson.ucla.edu, william.mann@anderson.ucla.edu. We thank Brent Ambrose, Zahi Ben-David, Morris Davis, Andrea Eisfeldt, Stuart Gabriel, Mark Garmaise, Eric Hughson, Ed Leamer, Lars Lochstoer, Hanno Lustig, Daniel McMillen, Tyler Muir, Stephen Oliner, Liang Peng, and conference and seminar participants at the AREUA meetings, the Minnesota Junior Finance Conference, the UCSD Law and Finance Conference, University of Mannheim, Goethe University in Frankfurt, the Frankfurt School of Finance, Hong Kong University, and Penn State for helpful comments. We gratefully acknowledge financial support from the UCLA Ziman Center for Real Estate Research and the Howard and Irene Levine Program in Housing and Social Responsibility.

# 1 Introduction

Housing is a large component of the wealth portfolio in the United States, both in the aggregate and for the typical household. While a growing literature studies the relationship between risk and return in housing markets (e.g., Han (2013) and Peng and Thibodeau (2013)), much remains unknown about the cross-sectional determinants of this relationship at a fine geographic level. In this paper, we establish that housing is a riskier investment for lower-income households. Our main finding is that lower-income zip codes experience higher housing return volatility without any higher average returns. We argue that the collateral role of housing, as emphasized by prior studies such as Lustig and Van Nieuwerburgh (2005), can explain these patterns in the data.

To frame our findings, we begin with a theory of housing return volatility. The representative household in our model is endowed with a risky income stream and is impatient relative to an exogenous borrowing rate. It thus borrows up to a limit that we impose via a standard collateral constraint. The price of housing therefore depends both on the future marginal rate of substitution (MRS) between housing and other consumption, and on the future value of the home. The collateral constraint means that the MRS fluctuates with income shocks, which produces volatility in the housing return as long as housing supply is not perfectly elastic. Moreover, this volatility is greater for lower-income households, for whom the magnitude of income shocks is relatively larger.

A decrease in the debt capacity of the house, via a tightening of the collateral constraint (e.g., a lowering of the maximum LTV ratio), causes the contemporaneous MRS to become a proportionally-larger component of the present value of the home as compared to its future price, amplifying the endogenous volatility in housing returns. Housing return volatility is further amplified if households have non-homothetic preferences over housing services, which causes the MRS to depend on their relative level. Both the LTV constraint and non-homothetic utility have been documented and studied in prior research on housing markets; we show that they combine to amplify endogenous volatility in housing returns.

Our model delivers several novel, testable predictions. Most importantly, higher average household income leads to lower volatility of housing returns in a given market, assuming that income volatility is decreasing in the level of income.[1] The model predicts this to be a smooth and robust relationship across time and space. Further, return volatility is also higher in the presence of tighter collateral constraints. Both predictions reflect the same mechanism: Collateral constraints prevent the market from smoothing out home prices relative to fluctuations in the household's MRS.

We corroborate our predictions empirically by measuring housing return volatility and income at the zip code level. We demonstrate that low-income zip codes feature consistently higher return volatilities, with no compensating increase in their average return. This finding holds across two data sources (CoreLogic and Zillow), and within each of the largest metropolitan statistical areas (MSAs) in the United States. In our main result, a doubling of annual income is associated with 1.3% less annual volatility in housing returns when measured with CoreLogic, or 2.7% less when measured with Zillow. Importantly, the *level* of housing returns is not any higher in the low-income, high-volatility zip codes.

We strengthen this conclusion by exploiting time-series variation *within* zip code: Increases in home values, which represent increases in household wealth, lead to lower housing return volatility. This relationship holds whether we look within state, MSA, or zip code. We also use the time-series dimension of the data to show that the income-volatility correlation is consistently negative throughout the sample period, although stronger in the post-crisis period, which arguably reflects a period of tighter constraints. Furthermore, we show that the results are not driven by correlation between income and liquidity, as turnover rates are roughly the same across bins of income throughout the sample period.

We also corroborate the second prediction of our model, that tighter collateral constraints lead to greater housing return volatility. To proxy for the tightness of collateral constraints, we measure the state-level degree of lender recourse, following the coding of Ghent and

[1] Guvenen, Karahan, Ozkan, and Song (2015) document that income volatility is highest for the lowest-income households, and is decreasing in income for 90% of the income distribution.

Kudlyak (2011). The prior research on lender non-recourse laws emphasizes that they constrain access to credit. In our model, a natural consequence of this effect is higher housing return volatility, and this indeed is what we find. Controlling for the effect of wealth, we find that states allowing a lesser degree of recourse also have greater return volatility, and that this finding is robust to demographic controls.

Aside from its direct effect on the household's portfolio problem, housing return volatility may also matter for the supply of housing. Real-option models of housing construction conclude that production of new homes is less frequent when housing demand is more volatile (e.g. Guthrie, 2010, Oh and Yoon, 2016). Our model induces demand volatility via financial constraints that matter more for low-income households, suggesting that housing supply may be less responsive to demand and price movements in lower-income areas.

To corroborate this prediction, we analyze new permit issuance and the age of the local housing stock using Census data. We show that both the level and growth rate of permit issuance are much more volatile in lower-income areas, while the housing stock is on average older in these areas. These both reflect less-frequent adjustments to the housing stock, which again is a natural implication of the endogenous volatility in demand induced by our model. Our findings provide a novel channel – financial constraints – by which housing supply may be particularly suboptimal in low-income areas, contrasting geographic constraints as studied by Saiz (2010) or inefficient regulation as discussed in Gyourko and Molloy (2014).

Our results are consistent with and add to earlier studies of housing return patterns. Ambrose, Buttimer, and Thibodeau (2001) find that the relationship between house price volatility is u-shaped in house price level for the Dallas metropolitan area. Peng and Thibodeau (2013) unconver the same negative relationship between housing return volatility and geographic income measures for the Denver metropolitan area that we find nationally. Peng and Thibodeau (2017) find that idiosyncratic house price risk is u-shaped in zip code level income; we study the total volatility rather than the idiosyncratic component of housing returns. Our results also complement those of Eisfeldt and Demers (2015), who present the first comprehensive look at the cross-section of rental housing returns for the entire US. Case, Cotter, and Gabriel (2011) find evidence for a single factor model of housing returns.

Our model of the housing market builds on the literature that emphasizes the importance of collateral constraints for asset markets. Kiyotaki and Moore (1997) show how the presence of collateral-constrained agents can amplify fundamental shocks in asset markets. Many studies have demonstrated the importance of this effect in real estate markets. For example Lamont and Stein (1999) and Almeida, Campello, and Liu (2006) demonstrate that house prices are more sensitive to shocks to economic fundamentals in locations in which households are more highly levered. More recently Justiniano, Primiceri, and Tambalotti (2015) study a model similar to Kiyotaki and Moore (1997) to show that collateral constraints can quantitatively explain many features of the housing boom and bust of the 2000's. Our model is similar in spirit to Justiniano et al. (2015), but bears closer resemblance to that of Rampini and Viswanathan (2010) and Rampini and Viswanathan (2013).

Another body of evidence shows that credit markets can have an important impact on house prices. Ben-David (2011) shows that financially constrained borrowers inflated house prices to draw larger mortgages. Ortalo-Magne and Rady (2006) highlight how young households' leverage in their first home can have an important effect on house price volatility. Landvoigt, Piazzesi, and Schneider (2015) present an assignment model of the housing market in which households face collateral constraints, and find that a key driver of variation of house prices within the San Diego metropolitan area was cheaper credit for poor households. Landvoigt (2017), in a quantitative model of housing markets featuring expectations and credit constraints, shows that an increase in price uncertainty rather than average expectations can explain the rise in household debt during the housing boom of the 2000's.

Housing as a source of collateral has also been shown to have important implications for the broader economy. Lustig and Van Nieuwerburgh (2005) show that a decrease in home values leads to a greater market price of risk, as collateral constraints make it harder for households to share risk. Mian and Sufi (2011) provide evidence that increased home equity during the early 2000's allowed for an increase in borrowing and the subsequent default crisis of the late 2000's.

To our knowledge, ours is the first model to integrate non-homethetic prefences into a dynamic model of house prices with collateral constraints. However, such preferences have been emphasized as an important driver of real estate markets. Notably, Albouy, Ehrlich, and Liu (2016) show that non-hometheticity can help explain the secular trend in housing expenditure shares.

# 2 Model

In this section, we present a model in which volatility arises in housing returns due to a representative household who is constrained, and therefore cannot perfectly smooth consumption, coupled with imperfectly elastic housing supply. We further show that this effect is amplified by non-homotheticity of utility over housing and non-housing consumption. The first purpose of the model is to demonstrate that a negative income-volatility correlation is expected as a robust equilibrium outcome, without appealing to government policies, microstructure issues, or historical anomalies as explanations. The second purpose is to show that salient features of housing markets amplify this relationship.

A representative household values consumption and housing according to the utility function $u(c_t, h_t)$, and discounts future consumption at rate $\beta$. The household is endowed with a risky income stream, which is the only fundamental source of uncertainty in the model, and we index its realizations by $s_t$. After realizing income $y_t$, the household repays state-contingent promises $\bar{b}(s_t)$ made last period; borrows a new amount $b_t$ via new repayment promises $\bar{b}(s_{t+1})$ for tomorrow; and purchases housing for next period $h_{t+1}$ at its current price $p_t$. Consumption is the residual between wealth, housing, and bond positions, and the price of the consumption good is normalized to 1.

Our model of collateralized lending is similar to that in Rampini and Viswanathan (2010), Justiniano et al. (2015), and Kiyotaki and Moore (1997): The lender discounts the household's promised repayments by a rate $R$. We shut down the equilibrium determination of this rate, treating it as exogenously determined by deep-pocketed agents who are outside of the model. We create demand for borrowing by assuming that the household is sufficiently impatient relative to that rate that he borrows as much as possible, that is we assume that $\beta << 1/R$. However, a collateral constraint limits this borrowing. The household can only promise to pay up to a fraction $\theta$ of the value of the house in any given state. This constraint can be motivated by assuming that loans are subject to limited enforcement.

The household's problem can be summarized as

$$\max_{c_t,h_t,b_t} \quad \mathbb{E}\left[\sum_t \beta^t u(c_t, h_t)\right] \tag{1}$$

$$s.t. \quad c_t + p_t h_t \leq W_t + b_t \tag{2}$$

$$W_{t+1}(s_{t+1}) \equiv y_{t+1}(s_{t+1}) - \bar{b}_t(s_{t+1}) + h_t p_{t+1}(s_{t+1}), \tag{3}$$

$$\bar{b}_t(s_{t+1}) \leq \theta h_t p_{t+1}(s_{t+1}), \tag{4}$$

$$b_t = \frac{\mathbb{E}[\bar{b}_t(s_{t+1})]}{R}. \tag{5}$$

Equations (2)-(3) jointly characterize the budget constraint, and equation (4) is the collateral constraint. Equation (5) is a lender optimality condition: The upfront loan proceeds $b$ are equal to the discounted value of those state-contingent promises.

We next impose our assumption that the household is impatient and will always borrow the maximum possible, so that the collateral constraint (4) always binds. This assumption is made for tractability, allowing us to study the real effects of the constraint without modeling the sources of interest rate fluctuations. It simplifies the problem in two steps: First, we set (4) to equality and substitute it into the definition of $W_{t+1}$ and into the final condition defining $b_t$. Second, we substitute that final condition into the RHS of the budget constraint.

With these simplifications, we can now formulate the household's problem recursively as

$$V(H_t) \equiv \max_{H_{t+1}} \quad u(c_t, H_t) + \beta \mathbb{E}[V(H_{t+1})] \tag{6}$$

$$\text{where} \quad c_t \equiv y_t + \theta \frac{1}{R} H_{t+1} \mathbb{E}[p_{t+1}] - \theta H_t p_t - p_t(H_{t+1} - H_t) \tag{7}$$

In the budget constraint, the first term is income; the second term is the amount of borrowing that can be done today against the future value of the home; the third term is repayment of borrowing that was made yesterday; and the last term is the cost of adjusting to the new level of housing.

This leads to an Euler equation for housing,

$$p_t = \mathbb{E}[M_{t+1} MRS_{t+1}] + \theta \times \frac{1}{R} \mathbb{E}\left[p_{t+1}\right] + (1-\theta) \times \mathbb{E}\left[M_{t+1} p_{t+1}\right]$$

where

$$M_{t+1} \equiv \beta \frac{u_1(c_{t+1}, H_{t+1})}{u_1(c_t, H_t)}$$

and

$$MRS_{t+1} \equiv \frac{u_2(c_{t+1}, H_{t+1})}{u_1(c_{t+1}, H_{t+1})}$$

The value of the home comes from two sources: First, tomorrow's utility over the flow of housing services (discounted by a standard pricing kernel); and second, the financial value of the home. This latter part is a weighted average of two sources. For the fraction $\theta$ against which the house borrows, the financial value of the house is its current debt capacity, given by the middle bracketed term. For the fraction $1-\theta$ against which the house does not borrow, the financial value is the future value of the home in the household's wealth portfolio. This is positive but, by assumption, strictly less than the value of being able to borrow against the home, so that the household always borrows the maximal fraction $\theta$.

We now introduce several more simplifications that allow us to analyze equilibrium in this model and illustrate the intuition of our argument:

First, we employ a particularly tractable specification of non-homothetic utility: additively separable, isoelastic utility over consumption and housing,

$$u(c,H) = \frac{c^{1-\alpha}}{1-\alpha} + \frac{H^{1-\gamma}}{1-\gamma}$$

The relative magnitudes of $\alpha$ and $\gamma$ govern the wealth effects: In particular, if $\gamma > \alpha$, as we will assume, then the consumption/housing ratio is decreasing in wealth. Also for tractability, we will consider the limiting case $\alpha \to 1$ in the household's utility function, so the non-housing component of this function is log utility. This maintains the wealth effect of housing consumption while greatly simplifying the solution by yielding tractable first-order conditions. We thus have $M_{t+1} = \beta \left(\frac{c_{t+1}}{c_t}\right)^{-1}$ and $MRS_{t+1} = c_{t+1} H_{t+1}^{-\gamma}$.

Second, we constrain the household to consume a fixed aggregate supply of housing $\bar{H}$, which abstracts away from the quantity decision in housing to focus on price and return dynamics, which are our main interest. This has several implications: Since we also set up the problem such that the bond position is always the maximal possible value, the budget constraint pins down the level of consumption every period at

$$c_t \equiv y_t + \theta \bar{H} \left( \frac{1}{R} \mathbb{E}[p_{t+1}] - p_t \right)$$

Conditional on a realization of $y$, this leaves only the prices $p$ as unknown equilibrium quantities, and implies that there are only as many states as realizations of $y$. It also means that the term $M_{t+1} MRS_{t+1}$ becomes deterministic (as of time $t$) and equal to $\beta c_t \bar{H}^{-\gamma}$.

Third, we assume that income $y$ follows a two-state Markov process, with probability $\pi$ of staying in a given state. Then we can replace $t$ in the Euler equation with $s \in \{H, L\}$ and decompose the expectation as

$$p_s = \beta c_s \bar{H}^{-\gamma} + \theta \times \frac{1}{R} \times (\pi p_s + (1-\pi) p_{-s}) + (1-\theta) \times \left( \pi p_s + (1-\pi) \frac{c_s}{c_{-s}} p_{-s} \right) \quad (8)$$

Given the unconditional probability 1/2 of being in either state, we can then compute the unconditional volatility of log returns:

$$Var\left(\ln \frac{p_{t+1}}{p_t}\right) = (1-\pi)(\ln p_H/p_L)^2$$

Volatility in this model depends on the price ratio between high and low states, and those are pinned down by the equilibrium conditions from before.

Although closed-form model solutions are not possible, we observe that equilibrium requires volatility in consumption, $c_H > c_L$. Intuitively, this happens because the only way the household can smooth consumption is if home prices fluctuate, as this shifts the always-binding constraint on the amount borrowed. If she could smooth consumption perfectly, all fluctuations in asset prices would disappear as $c$ and $H$ would both be constant.

To go further, we can numerically compute model solutions and comparative statics. Most importantly, Figures 1 and 2 summarize comparative statics on $y_L$, holding fixed the difference $y_H - y_L$ as well as all other parameters in the model. They illustrates the key wealth effect of our model: For example, following the dashed line in each figure, as we double the low-state income, the unconditional volatility of the housing falls roughly in half, from 2.6% to 1.6%. This corresponds with the correlations we will examine in the empirical section, and while the model is not designed for quantitative analysis, this effect is roughly the same order of magnitude as the results we will demonstrate.

Figures 1 and 2 also demonstrate that tightening the curvature of utility over housing (by increasing $\gamma$), and tightening the collateral constraint (by decreasing $\theta$), both significantly amplify the relationship between income and volatility. Intuitively, the first induces greater volatility in the MRS between housing and other goods, while the second causes that volatility to be a larger component of housing prices, by decreasing the value of future debt-capacity in the Euler equation. We view both as salient features of housing markets.

With these findings in place, we next examine how the data align with our model.

# 3 Data and measurement

We begin this section by demonstrating the key comparative static from the model: the volatility of housing returns is higher for lower-income households. We measure income and housing returns within Metropolitan Statistical Areas (MSAs), using finer zip-code-level geographic variation.[2]

To measure housing returns, and the volatility of those returns, we obtain home price data from two alternative sources. The first is the CoreLogic Single Family Combined Home Price Index (HPI), which is the standard in much real estate research. The second is the Zillow Home Value Index (ZHVI), a newer dataset. Our results are qualitatively similar with either index, but are more stark using the Zillow than the CoreLogic data. The primary difference between the two, which likely explains this discrepancy, is that CoreLogic is based on a repeat-sales methodology, capturing innovations to a home's value only when that home is actually sold. In contrast, Zillow's ZHVI uses hedonic regressions to update the value of *all* homes in a region in response to each transaction price.

For either the HPI or ZHVI, we use the time series for each zip code to construct two cross-sectional variables: the average 12-month return in home prices, and the standard deviation of that return. Specifically, for region $z$, we calculate

$$\begin{aligned}
r_{z,t}^{ann} &= 12 \times \ln\left(\frac{p_{z,t+1}}{p_{z,t}}\right), \\
\bar{r}_{z}^{ann} &= \frac{1}{T}\sum_{t=0}^{T} r_{z,t}^{ann}, \\
\bar{\sigma}_{z}^{ann} &= \sqrt{\frac{1}{T-1}\sum_{t=0}^{T}\left(r_{z,t}^{ann} - \bar{r}_{z}^{ann}\right)^2}.
\end{aligned}$$

where $t$ indexes months from January 1998 ($t = 0$) to October 2015 ($t = T$), and $p$ is the

[2]Strictly speaking, zip codes are not geographic concepts. Our references to zip codes are actually to Zip Code Tabulation Areas (ZCTAs), which are constructed by the US Census as geographic partitions of the United States that roughly correspond to actual zip codes.

zip-code-month level of the specific index employed.[3]

Finally, to measure household income, we obtain from the IRS zip-code-level statistics on Adjusted Gross Income (AGI) as reported in tax returns. These statistics are available at irregular frequencies beginning in 1998. For each zip code, the IRS reports both the number of tax returns and the total AGI across all returns, so we divide the two to obtain a mean AGI per household for each zip code. We use the 1998 cross-section of AGI, the earliest available, as the measure of household income throughout our analysis.

Our analysis is performed on the cross-section of 5,438 zip codes that have non-missing CoreLogic and Zillow indices for every month from January 1998 to October 2015. Figure 3 shows the distribution of average returns and return volatilies across these zip codes, comparing the numbers from the CoreLogic and the Zillow data. Figure 4 shows scatter plots of the CoreLogic and Zillow values against each other for a given zip code.

While the average housing return within a zip code calculated using either index appears roughly the same, the volatility of that return can be dramatically different between the two, with the Zillow volatilities typically lower and seeming to follow a skewed distribution, where the CoreLogic volatilities are higher on average and symmetrically distributed. The discrepancy in the estimated volatilities is intriguing, especially given that the estimated returns are so similar, but we simply use this as motivation to examine qualitative results based on both data sources rather than only one.

# 4 Volatility and income

We first observe that the volatility of housing returns has a very different cross-sectional distribution than the mean return. In particular, volatility is higher in lower-income areas. Figure 5 separates zip codes into six bins by AGI, and plots volatilities (in Panel (a)) and

[3]Note that we measure housing risk via total return volatility, which is consistent with our motivation of focusing on the household's portfolio problem, and with the view that households find it difficult to diversify their housing invevstments. Prior papers in this literature, following a different motivation, often focus instead on an idiosyncratic return component extracted from a statistical model of returns.

average returns (in Panel (b)). Volatility in Panel (a) is noticeably higher for lower-income areas. The spread in annual volatility between the lowest- and highest-AGI bins is roughly 2% in CoreLogic and over 3.5% in Zillow. This finding is consistent with the mechanism we present in our model and Guvenen et al. (2015), who find that household income volatility is decreasing in the level of income for 90% of the income distribution.[4] On the other hand, Panel (b) shows that this higher volatility is not compensated in the data by higher returns. If anything, returns seem to be slightly increasing in AGI, but there is no quantitatively meaningful relationship.

Figure 6 looks for the same pattern using only *within*-MSA variation. Returns and volatilities calculated with either index are adjusted by the MSA-level mean, and the six bins are recalculated separately for each MSA, so that they capture relative income position within-MSA instead of nationwide. Despite these adjustments, we see that the disparities in volatility across bins remains sizeable. Using the CoreLogic data, the highest AGI bin has 1.3% lower annual return volatility than the lowest-AGI bin, with no meaningful difference in annual return level. In the Zillow data, the disparity is larger, as before, at 2.88%. In both cases we see a steady decline in return volatility across the bins from low to high AGI.

Tables 1 and 2 display regressions confirming that these findings are statistically significant and robust, using the CoreLogic and Zillow data respectively. Instead of bins of AGI, the logarithm of zip-code mean AGI is used as the independent variable in the regressions, and all regressions include MSA fixed effects to preserve the within-MSA interpretation of Figure 6. Standard errors are clustered by state to allow for possible geographical clustering in the residuals.

Column (1) of each table performs this regression in the full sample of zip codes. The estimated coefficients suggest that, within-MSA, a doubling of income is associated with 1.2% lower annual housing return volatility as measured through CoreLogic data, or with

[4] Guvenen et al. (2015) do find that income volatility increases with income for the top 10% of the income distribution at the household level. However, our data are at the zip code level, and this aggregation is likely to mask the effects of the top of the income distribution.

2.7% less annual volatility as measured through Zillow. Column (2) of each table shows that this result is not driven by relatively sparsely-populated MSAs; if anything, the estimated effects strengthen slightly when the analysis is restricted to MSAs with at least a million 1998 tax returns. This reduces, by more than half, the number of zip codes in the regression, but the coefficients remain statistically significant. Meanwhile, Columns (3) and (4) reiterate that the higher volatility of housing returns in low-income zip codes is not associated with higher average returns; if anything, the relationship is slightly in the opposite direction.

Figure 7 demonstrates these findings visually with zip-code maps of three of the largest MSAs in the sample, Los Angeles (Panel (a)), Chicago (Panel (b)), and Atlanta (Panel (c)). These figures measure housing returns with the Zillow index. For all three panels, the left figure shades zip codes according to eight bins of the volatility of the Zillow HVI return from 1998-2015, with darker shading corresponding to more volatility. In Los Angeles, for example, the returns to housing have been most volatile in poorer areas to the south and in the San Fernando valley. The right figure in each panel shades zip codes according to eight bins of 1998 mean AGI, but with darker shading corresponding to lower income. The resemblance to the left figures in each panel is striking. Put simply, high-volatility zip codes are also low-income zip codes.

Similar figures to Figure 7 can be constructed for every major MSA in the country (available on request). To summarize the figures, Tables 3 and 4 perform the prior within-MSA regression of return volatility on log AGI, explicitly breaking out each of the 16 largest MSAs in the sample, and using (respectively) the CoreLogic index and the Zillow index to measure housing returns. In all 32 specifications, the point-estimate of the coefficient on log AGI is negative. It is economically large in most, and statistically significant in all but two.[5]

[5] The effect we find in Denver is consistent with Peng and Thibodeau (2017) (e.g., their Table 5), although they extract idiosyncratic return volatility from a statistical model, whereas we focus on total return volatility.

## 4.1 Longer holding period returns

Our main results use the volatility of the monthly housing price return, where this return is annualized simply by scaling up by a factor of 12. For our effects to be driven by credit constraints, the relevant return frequency should be the horizon over which the household is unable to finance consumption without borrowing. Arguably, one month is too short a horizon for this. If the housing price return process is i.i.d., this makes no difference to the results, because variance would scale linearly as we calculate. But if returns are negatively autocorrelated over a horizon of a few months, we may overstate housing return volatility by using a monthly return series. Intuitively, households could wait out a negative, but transitory, shock to home prices, simply by borrowing a month or two later.

In this section, we show that this issue is not driving our results. To do this, we recalculate the housing return and volatility series using year-over-year housing returns instead of monthly returns. That is, we define

$$
\begin{aligned}
r_{z,t}^{yoy} &= \ln\left(\frac{p_{z,t+12}}{p_{z,t}}\right), \\
\bar{r}_{z}^{yoy} &= \frac{1}{T}\sum_{t=0}^{T} r_{z,t}^{yoy}, \\
\bar{\sigma}_{z}^{yoy} &= \sqrt{\frac{1}{T-1}\sum_{t=0}^{T}\left(r_{z,t}^{yoy} - \bar{r}_{z}^{yoy}\right)^{2}}.
\end{aligned}
$$

The only difference from the main specification is the replacement of monthly returns with year-over-year returns in the first line.

Tables 5 and 6 repeat the regressions of Figures 1 and 2, using this year-over-year specification. We retain only January of each year from the panel of zip codes, then calculate means and standard deviations of the January-to-January home price appreciation for each zip code in the panel. The results are economically and statistically indistinguishable from the earlier tables.

## 4.2 Time-series stability

We next investigate the time-series stability of the negative within-MSA income-volatility correlation. Of particular interest is the evolution of this correlation around the years of the financial crisis, which occurs roughly midway through our sample. We repeat our main specification, a cross-sectional regression at the zip-code level with MSA fixed effects, using rolling five-year windows instead of the whole time series, and clustering all standard errors by MSA. Figure 8 plots the time series of resulting regression coefficients and their standard errors against the first year in the rolling window.

Figure 8(a) shows that the negative cross-sectional relationship between income and return volatility is present, highly stable, and statistically different from zero, in both the Zillow and CoreLogic data, for the entire time series from 1996-2015. As with the main findings above, the magnitude of the income-volatility relationship is uniformly stronger for the Zillow data throughout the time series: In the early years of the time series, a log-point income increase is associated with about 35bp lower log return volatility in CoreLogic, compared to about 120bp in Zillow.

Interestingly, both correlations become significantly stronger towards the end of the sample, with the above magnitudes increasing to about 140bp in CoreLogic and to about 300bp in Zillow in regressions including years 2007 or later. This may demonstrate that the mechanisms for the negative correlation in our model have gotten stronger in the years since the crisis. However, this effect is not driven by just the crisis years: It is present and stable in both the earliest and latest regression windows, which exclude those years. Most importantly, the magnitude is stable, always negative, and always statistically significant.

## 4.3 Liquidity and income

We next address one potential explanation for the income-volatility correlation: cross-sectional variation in liquidity of housing markets. One might hypothesize that low-income housing markets feature lower liquidity as well, and therefore that the less-frequent updating of home price indices in these areas leads to the appearance of higher volatility in realized prices, even though the potential selling price of a house is not more volatile. Figure 9 addresses this hypothesis by examining zip-code level turnover rates (available from Zillow) across five bins of 1998 income and over the time series.

While the figure reveals substantial time-series variation in turnover, rising to over 8% in 2005 and then falling back to 5%, there is far less cross-sectional variation: The largest difference between the highest and lowest income bins is roughly a percentage point, and this occurs in the early part of the sample. During the peak turnover years, there is hardly any difference across income bins, and in 2009-2010, when the income-volatility correlation was at its strongest (see Figure 8), a small disparity briefly appears in the *opposite* direction of what we should expect: The lowest-income bin exhibits the highest turnover and thus liquidity. We conclude that a liquidity-income relationship cannot account for our documented volatility-income relationship.

## 4.4 Panel evidence

In this section, we exploit the panel dimension of the zip-code-month panel to further investigate our mechanism. Our goal is to show that volatility responds to *changes* in wealth in the direction one would expect based on our model. This exercise helps isolate our proposed mechanism from several alternative interpretations, most importantly any omitted variables that are fixed in the cross-section or that do not fluctuate with wealth.

Our instrument for household wealth is the lagged return of either of the two home value indices. Intuitively, an individual observation of this high-frequency (monthly) return has a persistent effect on the wealth of homeowners, and our model predicts that this wealth effect should then alter the volatility of future housing returns. On the other hand, outside of our proposed mechanism, there is no obvious reason for individual monthly returns to have persistent effects on volatility. Thus, if high (low) individual monthly housing returns predict low (high) future *volatilities* of monthly returns within the same zip code, we will regard this as evidence of our proposed mechanism at work.

To implement this logic, we calculate rolling volatility measures at the zip code level for both of our indices based on the prior 12 months of returns, starting in 1990. We regress this rolling return volatility on lags of the monthly return series. To avoid using observations based on overlapping windows, we retain only January of every year in these regressions, so the regression is performed on a zip-code-year panel, and we lag the returns on the right-hand side by a year or more. Our results are presented in Tables 7 and 8.

Table 7 shows that housing returns negatively predict future volatility within a zip code. Panel (a) uses the CoreLogic HPI series to measure housing returns and volatilities, and Panel (b) uses the Zillow ZHVI series. In both cases, the coefficient on distributed lags of the monthly housing return is significant and negative. The interpretation is that a positive (negative) wealth shock, via a positive (negative) individual monthly housing return, predicts a lower (higher) future degree of volatility in housing returns.

The magnitudes are sizeable: In the first column of Panel (a), a one-standard-deviation increase in the HPI return a year ago predicts a 0.13 standard deviation decrease in current volatility of the HPI return, based on sample standard deviations of 0.0156 and 0.0045 respectively. Moreover, the dynamics of the effects decay at longer lags, which is intuitive. Columns 2 through 4 show that the magnitudes of the coefficients are virtually unchanged when including state, MSA, and finally zip-code fixed effects. In Panel (b), the Zillow series exhibits the same qualitative effects, although the magnitudes are smaller: A one-standard-deviation increase in ZHVI return a year ago predicts a 0.04 standard deviation decrease in current volatility, based on sample standard deviations of 0.0097 and 0.0038 respectively.

We can pin down our interpretation even further by exploiting cross-sectional variation in household income, as in the previous section. A change in housing value should have a proportionally bigger effect among households that are poorer to begin with, and therefore should lead to a larger subsequent effect on volatility. Thus, we expect that the magnitude of the coefficients from Table 7 should be relatively higher in relatively low-income zip codes.

Indeed, Table 8 documents exactly this relationship. This table repeats the analysis of Table 7, interacting all explanatory variables with 1998 log AGI (the same measure of income employed in the previous section), after demeaning that variable across the full sample. (For compactness, only three lags are employed instead of four).

For a household of average income, Table 8 continues to document the negative relationship between lagged return and future volatility as in Table 7. However, a significant and positive coefficient on the interaction with income indicates that the relationship is stronger (weaker) for lower (higher) income households. With the CoreLogic data, this interaction is not significant beyond a one-year lag, but with the Zillow data it shows up two years later. Again, our interpretation is that the wealth effect of a monthly housing return is proportionally larger in areas with lower income (which proxies for lower wealth). Our model then predicts that the relationship between housing return and future volatility is stronger in lower-income areas, a prediction that is confirmed in the data.

# 5 Housing expenditure share and non-recourse status

Income has the advantage of being available at a finely-disaggregated geographic level, but it does not directly measure the fundamental forces in the model. In this section we employ cross-sectional predictors that are more coarsely aggregated, but may come theoretically closer to capturing the core ingredients of our model. We return to the cross-section of zip codes that was used in most of the prior results.

First, we seek a proxy for wealth effects in utility. When households have non-homothetic preferences over housing, income effects cause the housing expenditure share to fall as wealth increases (see Albouy et al., 2016). To capture this empirically, we obtain data on the housing expenditure share. Our data source is the Metropolitan Statistical Area Tables from the Consumer Expenditure Survey conducted by the Bureau of Labor Statistics, which provides characteristic spending patterns of households in many of the largest MSAs.

Through the 2003-2004 vintage, the survey covered 28 different MSAs, but thereafter it was decreased to 18. We use the 2003-2004 vintage to have the largest and most recent data possible. The dataset reports the MSA-level average expenditure share on a variety of goods, including housing. Our key variable of interest is total expenditures on housing, divided by the household's total annual expenditures. This variable falls within a tight range, between 0.3 and 0.4 for all 28 MSAs in the sample.

We next seek out variation in $\theta$. A commonly-explored source of such variation is the degree of lender recourse in the case of default, which varies substantially around the country due to state laws adopted in past decades, mostly in response to idiosyncratic circumstances (see Ghent (2014)). In some states, such as Florida, lenders can pursue a deficiency judgment granting a claim on the borrower's other assets; while in others, such as California, the lender must be satisfied with foreclosure and sale of the house itself. Ex ante, this should limit the amount the household can borrow as a fraction of its home value, leading in our model to a lower $\theta$, a tighter collateral constraint, and greater volatility in home price appreciation.

To classify states in terms of non-recourse status, we employ the coding of Ghent and Kudlyak (2011), who conduct a detailed reading of state-level policies banning or hindering deficiency judgments against residential properties. Under their coding, California, Washington, North Carolina, Arizona, Minnesota, Wisconsin, Oregon, and Iowa are coded as non-recourse states.[6]

We find that a higher MSA-level housing expenditure share, and the state-level indicator for non-recourse status, are both associated with higher housing return volatility. The results are presented in Table 9. Column 1 regresses return volatility on both of these new predictors: The coefficient on housing expenditure share is 0.712, indicating that a household that spends ten more percentage points of its total expenditures on housing experiences 7 percentage points higher housing return volatility on average. Meanwhile, a household in a non-recourse

[6] In 2009, in the middle of our sample period, Nevada also passed an anti-deficiency law. See Li and Oswald (2014) and Ghent (2014). We code Nevada as a recourse state, as in Ghent and Kudlyak (2011), but our results are not meaningfully affected if we instead drop Nevada.

state experiences a 3.4 percentage point higher housing return volatility. Both of these are compared to a cross-sectional standard deviation of 4.4 percentage points. To isolate this effect from the previous income results, and from variation in the sheer size of a zip code, Column 2 adds our other wealth proxy, the zip-code mean AGI, and Column 3 adds the zip-code number of tax returns filed with the IRS.

To explore the effect of non-recourse status further, we observe that its effect in our model should come through a reduction in realized credit, for which loan-to-value (LTV) is a good proxy. We obtain from CoreLogic the zip-code level median LTV ratio, average this value for each zip code throughout the sample period 1998-2015, and employ this average as an outcome variable in Column 5. Non-recourse states have LTVs that are about 2 percentage points lower on average, which is about one-third of the cross-sectional standard deviation. This finding does not appear to be present in the literature, and substantiates the idea that recourse status is important for credit availability.

Having shown that non-recourse status affects credit, which in turn affects volatility, a natural step is to combine these effects in an instrumental-variables (IV) regression, translating their magnitudes into a marginal effect of an increased LTV on return volatility. Column 6 reports this IV regression. If one accepts the exclusion restriction that a state's non-recourse status only affects housing return volatility through limits on credit availability, the estimated coefficient on LTV (equal to the ratio of the coefficients in Columns 1 and 5) has a direct interpretation in terms of our model: A percentage point increase in $\theta$ would lower return volatility by 1.4 percentage points.

# 6 Income and housing supply

Endogenous volatility in demand for and returns to housing is clearly important for the household's savings and consumption decisions, but it can have other real effects too. Most importantly, in a standard real-option model of home construction, volatility in home values increases the value of the option to delay. For example, Guthrie (2010) investigates such a model, and shows that an increase in demand volatility leads to less-frequent but more-intense development of new homes (see his page 66). Oh and Yoon (2016) also examine quantitative implications of a real-options model of housing production.

Our model produces endogenous volatility in housing demand through a combination of financial constraints and non-homothetic utility. This potentially provides a novel explanation for why housing supply may exhibit less-frequent adjustments in low-income areas. Media reports and anecdotal evidence suggest that this pattern is real and leads to affordability problems. In this section, we provide novel evidence of this pattern using county-year Census data on the issuance of building permits.

Table 10 runs regressions of various moments of the county-level time series on county-level mean AGI. Columns 1 and 2 show that high-income areas see more permits issued per year on average, but there is no significant relationship between income and the growth rate of permit issuance during our sample period. However, Columns 3 and 4 show that the *volatility* of permit issuance, in level and growth rate respectively, is strongly decreasing in income. Column 4 corresponds with the relationship depicted in Figure 10. All standard errors are clustered by state. Panel (b) of the table repeats all regressions with state fixed effects (the MSA fixed effects of the earlier analysis are not possible, since the permits data are only available at the county level).

Figure 10 summarizes the last of these regressions visually. The horizontal axis plots 1998 county-level mean AGI. The vertical axis plots the volatility of the log growth rate of permits issued for single-unit buildings from 1998-2015. We observe a strongly and monotonically negative relationship between the two variables. Our interpretation of these results is that volatility in the growth rate of permits reflects less-frequent but more-intense responses of supply to demand shocks.

As another way of thinking about this mechanism, Table 11 analyzes the local age of the housing stock, also reported by the Census (in the housing characteristics component of the most recent Public Use Microdata Samples from the American Community Survey). The dependent variable is either the age of the property as of 2015, or an indicator for the property being built in 2005 or later.[7] The geographic unit here is the Public Use Microdata Area (PUMA), the smallest geographic unit available in the ACS, as of 2000.

The results in Table 11 show a clear negative relationship between household income and the age of the housing stock. In columns 1 and 2, doubling a household's income is associated with a 2-3 year decrease in the age of the property, or a 2% greater probability that the house was built after 2005, in a regression with state fixed effects. Columns 3 and 4 report similar results using tighter fixed effects for the PUMA. Finally, Columns 5 and 6 take a nonparametric approach by constructing five bins of income within-state, showing that the relationship between income and age of the housing stock is monotonic.

Our interpretation of these findings is that construction activity is less frequent in low-income areas, consistent with the implications of higher housing return volatility in a real-options model of home construction. Existing explanations for low housing supply or sluggish adjustment focus on supply constraints arising from geography (Saiz, 2010) or regulation (Gyourko and Molloy, 2014). In contrast, our mechanism operates through financial constraints on the demand side, providing a potentially more-fundamental explanation for a lack of affordable housing supply.

# 7 Conclusion

In this paper, we demonstrate empirically and theoretically that two widely-studied features of housing – its collateral value for constrained households, and the non-homotheticity of preferences over it – lead in equilibrium to greater volatility of home price appreciation for low-income households, without any compensating increase in expected return. Our theoretical analysis assumes no frictions in mortgage markets beyond the limited borrowing

[7] Before 2004, the age of the property is only reported within a range: 2000-2004; or the decade in which the property was built if before 2000. We use the midpoint of the age range in each case. Also, properties built before 1940 are bottom-coded into one category; we drop these from the sample.

capacity of the impatient household. In fact, the model could likely be applied to a wide range of durable goods, although housing is its natural setting. Likewise, our empirical analysis did not focus on any particular time period (such as the housing boom or bust) nor on any particular region. Our results thus capture a fundamental connection between financial constraints and the return patterns of assets with collateral value in the presence of non-homothetic preferences. Because housing is such a large fraction of expenditures for the typical household, this is a quantitatively important pattern to understand for policy issues, such as problems with housing affordability.

# References

Albouy, D., Ehrlich, G., Liu, Y., 2016. Housing demand, cost-of-living inequality, and the affordability crisis. Working paper.

Almeida, H., Campello, M., Liu, C., 2006. The financial accelerator: evidence from international housing markets. Review of Finance 10 (3), 321–352.

Ambrose, B. W., Buttimer, R., Thibodeau, T., 2001. A new spin on the jumbo/conforming loan rate differential. The Journal of Real Estate Finance and Economics 23 (3), 309–335.

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

Case, K., Cotter, J., Gabriel, S., 2011. Housing risk and return: Evidence from a housing asset-pricing model. The Journal of Portfolio Management.

Eisfeldt, A., Demers, A., 2015. Rental yields and hpa: The returns to single family rentals. Tech. rep., National Bureau of Economic Research.

Ghent, A., 2014. How do case law and statute differ? Lessons from the evolution of mortage law. Journal of Law and Economics 57 (4), 1085–1122.

Ghent, A., Kudlyak, M., 2011. Recourse and residential mortgage default: Evidence from us states. Review of Financial Studies 24 (9), 3139–3186.

Guthrie, G., 2010. House prices, development costs, and the value of waiting. Journal of Urban Economics 68, 56–71.

Guvenen, F., Karahan, F., Ozkan, S., Song, J., January 2015. What do data on millions of u.s. workers reveal about life-cycle earnings risk? Working Paper 20913, National Bureau of Economic Research.
URL `http://www.nber.org/papers/w20913`

Gyourko, J., Molloy, R., 2014. Regulation and housing supply.

Han, L., 2013. Understanding the puzzling risk-return relationship for housing. Review of Financial Studies 26 (4), 877–928.

Justiniano, A., Primiceri, G. E., Tambalotti, A., 2015. Credit supply and the housing boom. NBER Working Paper No. 20874.

Kiyotaki, N., Moore, J., 1997. Credit Cycles. The Journal of Political Economy 105 (2), 211–248.

Lamont, O., Stein, J. C., 1999. Leverage and house-price dynamics in us cities. RAND Journal of Economics 30 (3), 498–514.

Landvoigt, T., 2017. Housing demand during the boom: The role of expectations and credit constraints. The Review of Financial Studies 30 (6), 1865–1902.

Landvoigt, T., Piazzesi, M., Schneider, M., 2015. The housing market (s) of san diego. The American Economic Review 105 (4), 1371–1407.

Li, W., Oswald, F., 2014. Recourse and residential mortgages: The case of nevada. Working paper.

Lustig, H. N., Van Nieuwerburgh, S. G., 2005. Housing collateral, consumption insurance, and risk premia: An empirical perspective. The Journal of Finance 60 (3), 1167–1219.

Mian, A., Sufi, A., 2011. House prices, home equity-based borrowing, and the us household leverage crisis. The American Economic Review 101 (5), 2132–56.

Oh, H., Yoon, C., 2016. Time to build and the real-options channel of housing supply. Working paper.

Ortalo-Magne, F., Rady, S., 2006. Housing market dynamics: On the contribution of income shocks and credit constraints. The Review of Economic Studies 73 (2), 459–485.

Peng, L., Thibodeau, T., 2013. Risk segmentation of american homes: Evidence from Denver. Real Estate Economics 41 (3), 569–599.

Peng, L., Thibodeau, T. G., 2017. Idiosyncratic risk of house prices: Evidence from 26 million home sales. Real Estate Economics 45 (2), 340–375.

Rampini, A. A., Viswanathan, S., 2010. Collateral, risk management, and the distribution of debt capacity. The Journal of Finance 65 (6), 2293–2322.

Rampini, A. A., Viswanathan, S., 2013. Collateral and capital structure. Journal of Financial Economics 109 (2), 466–492.

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

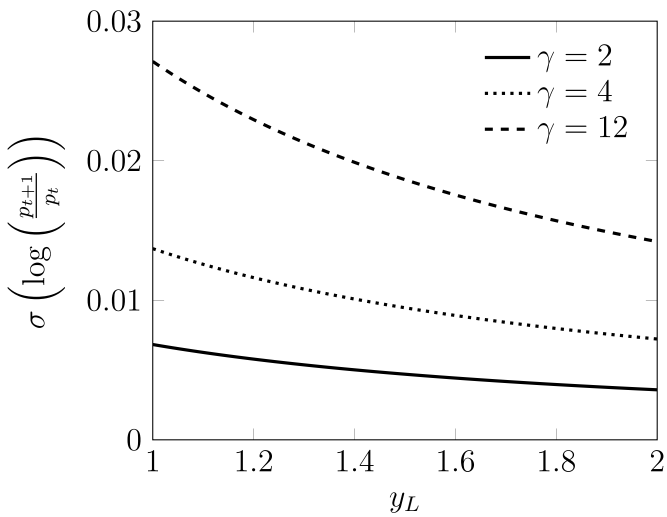



Figure 1: Comparative statics of return volatility with respect to $y_L$, holding fixed $y_H - y_L = 0.2$, for various values of $\gamma$. Other parameter values are: $\bar{H} = 2$; $\theta = .8$; $\pi = 0.52$; $R = 1.02$; $\beta = 0.9$. The model is solved numerically at each set of values.

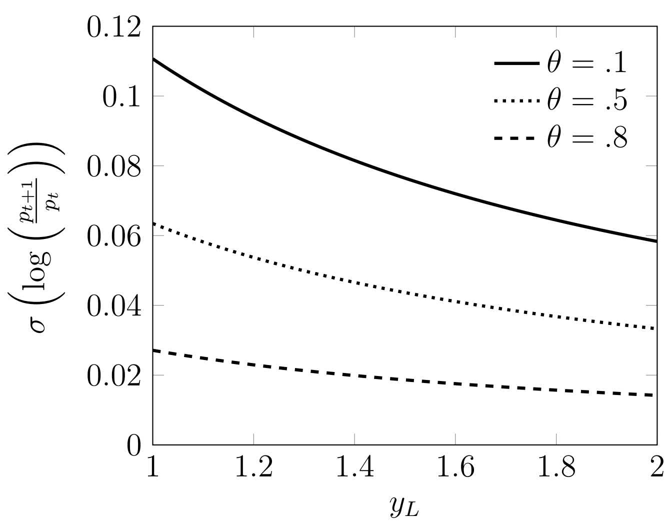



Figure 2: Comparative statics of return volatility with respect to $y_L$, holding fixed $y_H - y_L = 0.2$, for various values of $\theta$. Other parameter values are: $\bar{H} = 2$; $\gamma = 12$; $\pi = 0.52$; $R = 1.02$; $\beta = 0.9$. The model is solved numerically at each set of values.

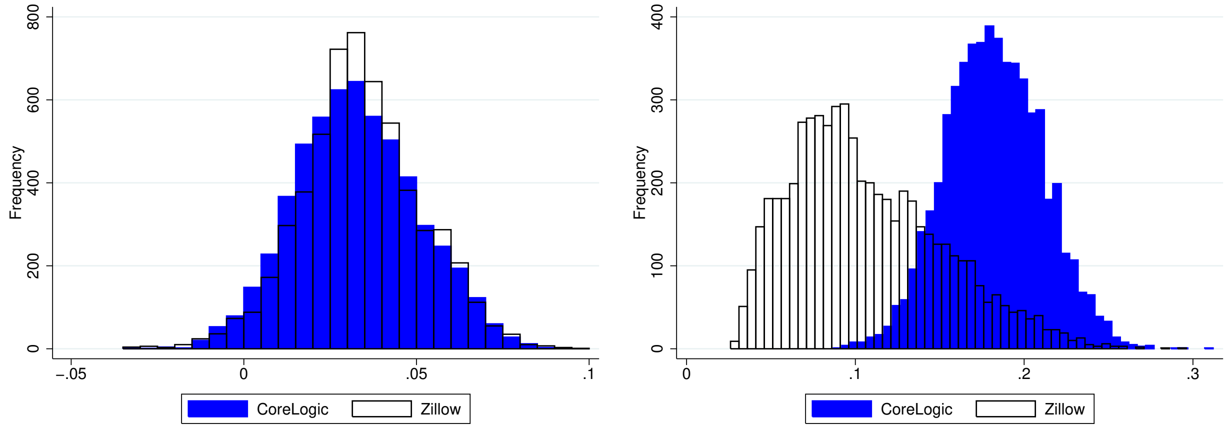



Figure 3: Distribution of zip-code level averages of annualized log monthly returns (left panel) and volatilities (right panel), 1998-2015. The solid blue bars are calculated using the CoreLogic Home Price Index, Single Family Combined series. The black outlined bars are calculated using the Zillow Home Value Index.

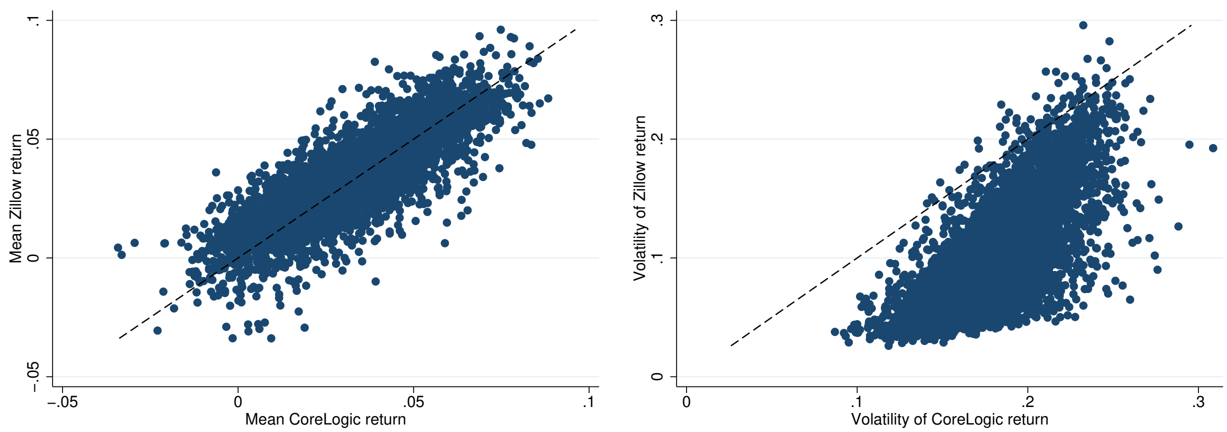
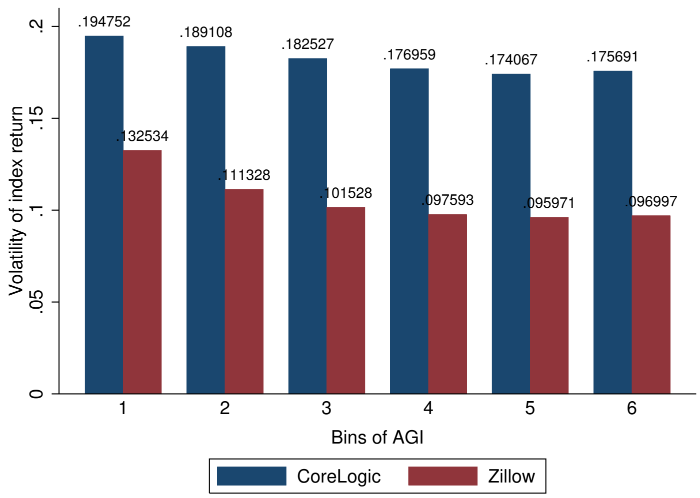
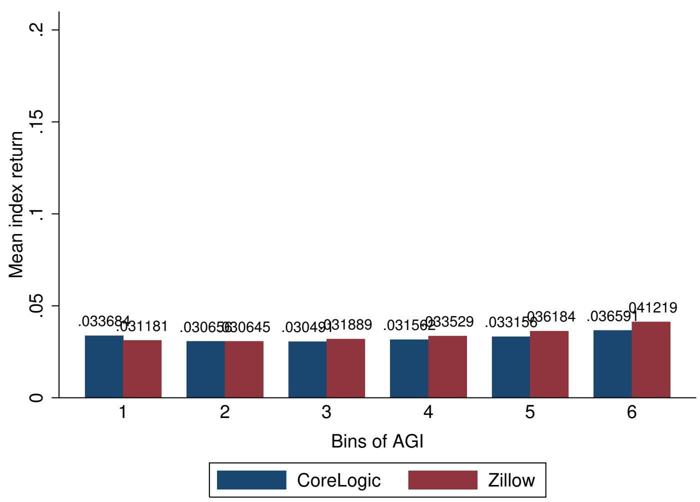



Figure 4: Scatter plot of zip-code level average returns (left panel) and volatilies (right panel) calculated using Zillow data against those calculated using CoreLogic data.

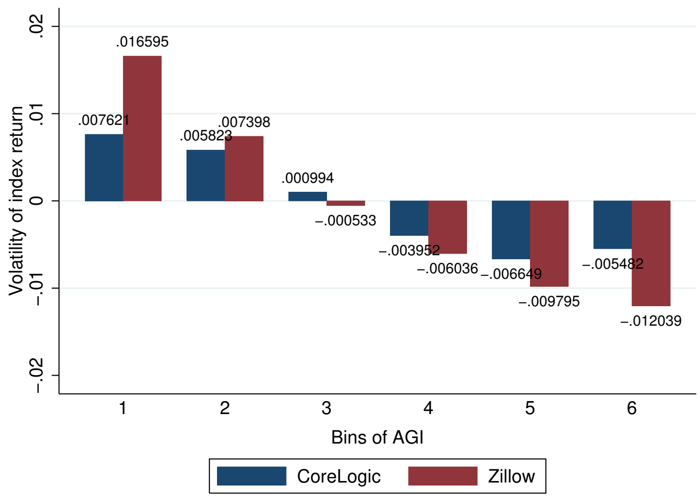



(a) Zip-code level housing return volatility, 1998-2015, by bins of 1998 mean adjusted gross income (AGI). Blue bars are calculated using CoreLogic data, and red bars using Zillow data.

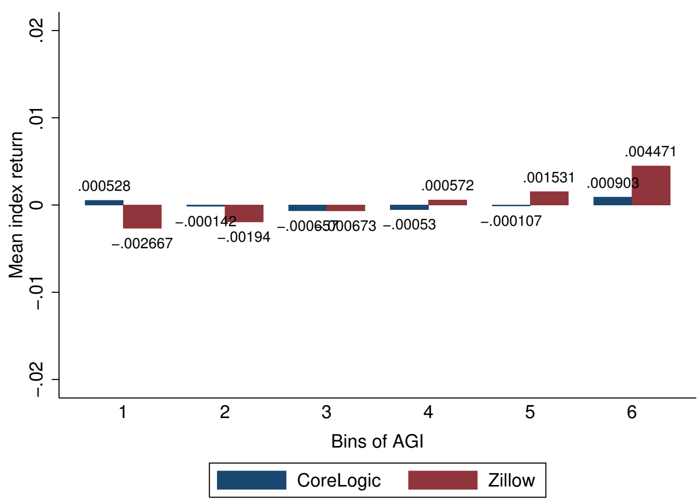



(b) Zip-code level average housing return, 1998-2015, by bins of 1998 zip code-level mean adjusted gross income (AGI). Blue bars are calculated using CoreLogic data, and red bars using Zillow data.

Figure 5: Housing returns and volatilities across zip codes.

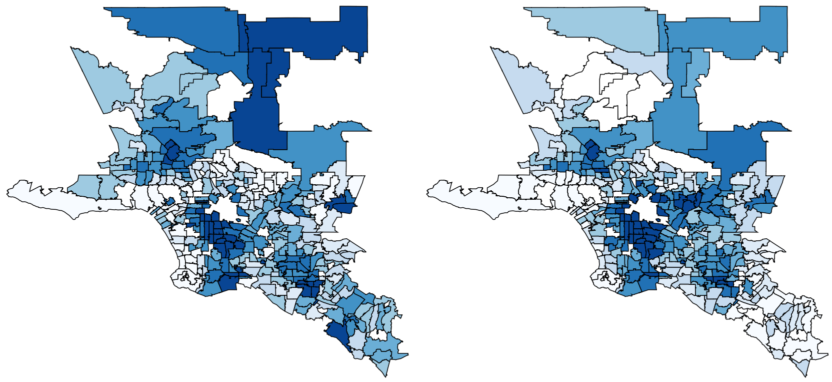



(a) Zip-code level housing return volatility, 1998-2015, by bins of 1998 zip code-level mean adjusted gross income (AGI). Blue bars are calculated using CoreLogic data, and red bars using Zillow data. Volatilities are demeaned within-MSA, and the bins are also constructed within-MSA.

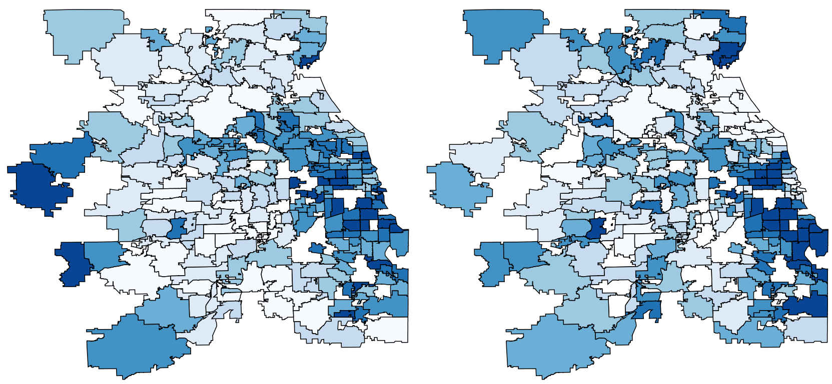



(b) Zip-code level average housing return, 1998-2015, by bins of 1998 zip code-level mean adjusted gross income (AGI). Blue bars are calculated using CoreLogic data, and red bars using Zillow data. Returns are demeaned within-MSA, and the bins are also constructed within-MSA.

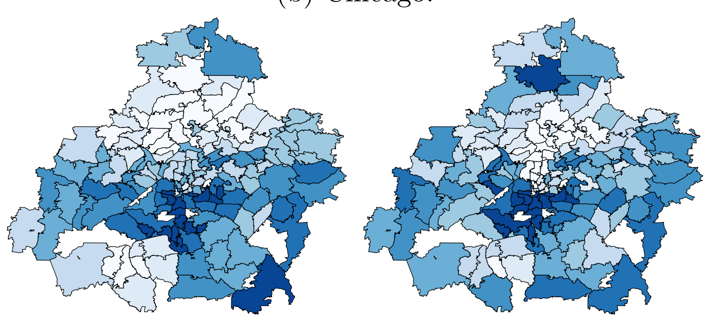
Figure 6: Housing returns and volatilities across zip codes, within-MSA.



(a) Los Angeles.



(b) Chicago.



(c) Atlanta.

Figure 7: Income and housing return volatility for three major MSAs. The three panels on the left show the annualized zip-code level volatility of home price returns from 1998-2015, based on Zillow data. Darker shading corresponds to higher volatility, using eight bins. The three panels on the right show zip-code level 1998 adjusted gross income, again using eight bins, but here darker shading corresponds to lower AGI.

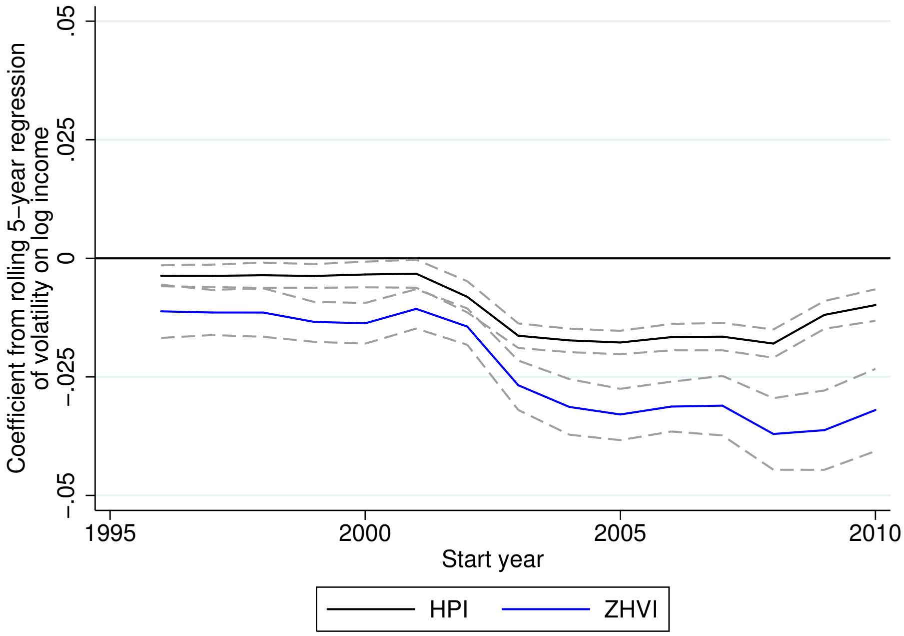



Figure 8: Coefficients from zip-code level regressions of return volatility and level on 1998 AGI, using rolling five-year windows starting at the year indicated, and including MSA fixed effects and clustering standard errors by state.

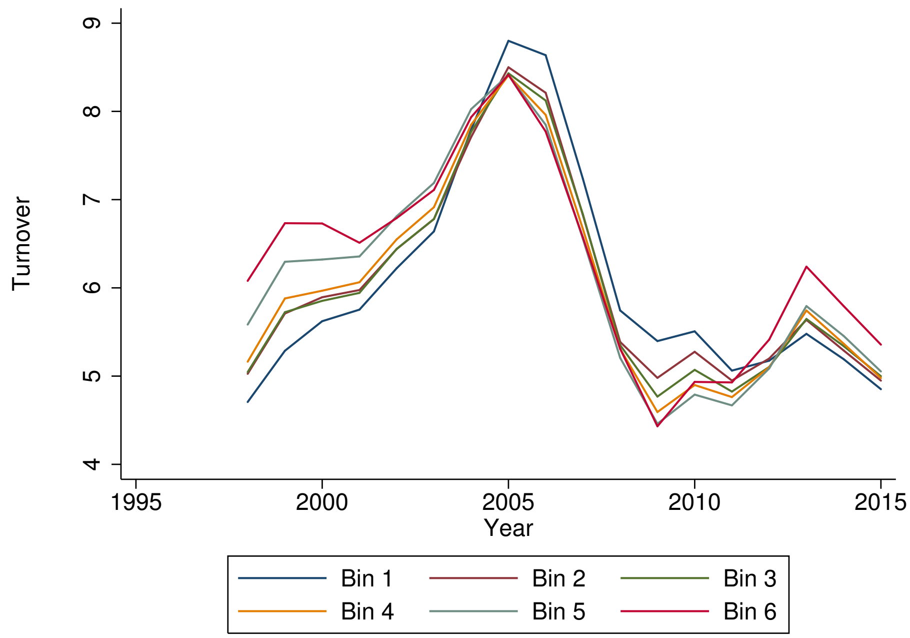



Figure 9: Fraction of properties turning over in a given zip code and month, by five bins of income. Source: Zillow and IRS.

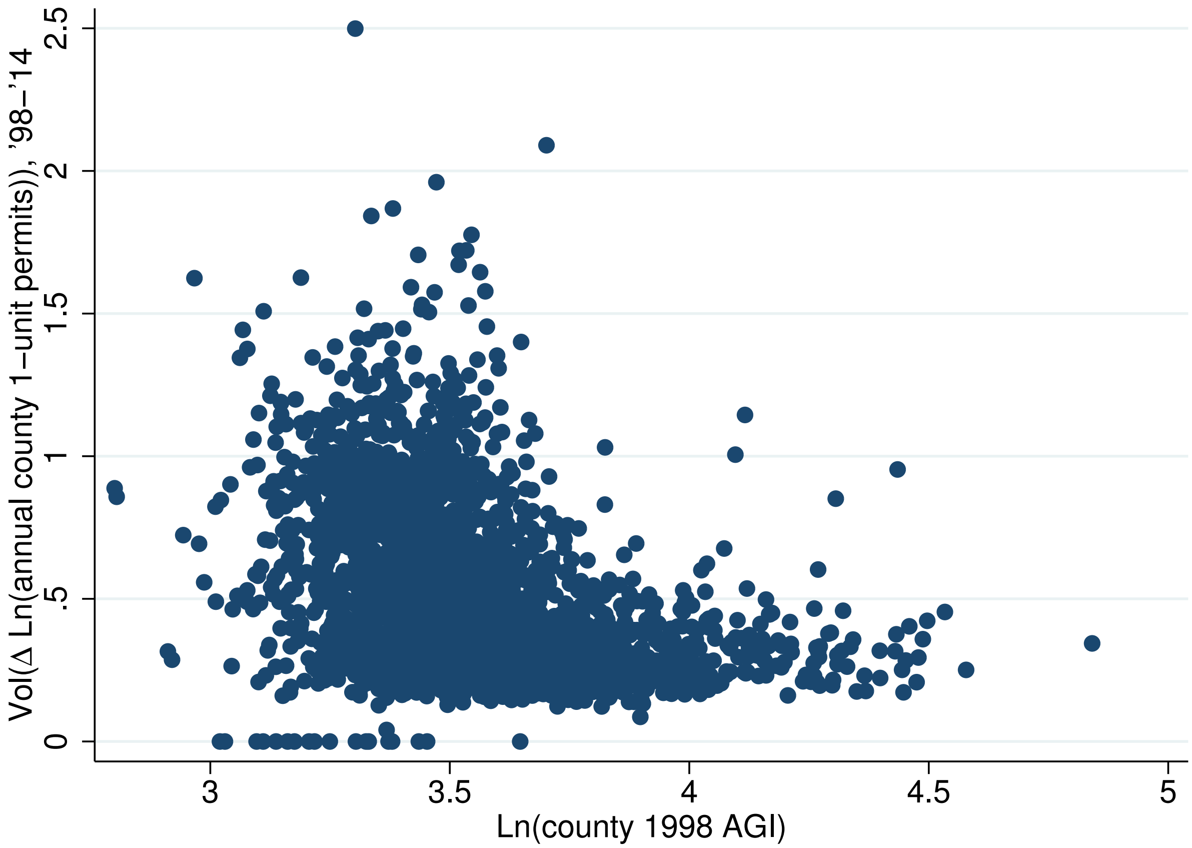



Figure 10: County-level data: Volatility of the annual log growth in permits issued for single-unit buildings, 1998-2015, versus log of 1998 mean AGI per return.

| | (1) HPI return vol | (2) HPI return vol | (3) HPI return | (4) HPI return |
|---|---|---|---|---|
| Ln(Mean AGI) | -0.0128*** | -0.0127*** | -0.0000430 | -0.00140 |
| | (0.00136) | (0.00165) | (0.000989) | (0.00141) |
| Fixed effect | MSA | MSA | MSA | MSA |
| Obs. | 5438 | 2312 | 5438 | 2312 |
| $R^2$ | 0.0940 | 0.113 | 0.00000350 | 0.00394 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1: In the first two columns, the dependent variable is $\bar{\sigma}_z^{ann}$, the volatility of the zip-code-level annualized monthly log housing return. In the last two columns, the dependent variable is $\bar{r}_z^{ann}$, the average of that return. Returns are measured using the CoreLogic Home Price Index (Single Family Combined) for a cross-section of 5,573 zip codes from 1998-2015. The explanatory variable is the natural logarithm of the zip code's mean adjusted gross income (AGI) from 1998, as reported by the IRS. All regressions include MSA fixed effects. Standard errors are clustered by state. Columns (2) and (4) restrict to MSAs in which one million or more tax returns were filed with the IRS in 1998.

| | (1) ZHVI return vol | (2) ZHVI return vol | (3) ZHVI return | (4) ZHVI return |
|---|---|---|---|---|
| Ln(Mean AGI) | -0.0272*** | -0.0309*** | 0.00630*** | 0.00533** |
| | (0.00278) | (0.00282) | (0.00143) | (0.00201) |
| Fixed effect | MSA | MSA | MSA | MSA |
| Obs. | 5438 | 2312 | 5438 | 2312 |
| $R^2$ | 0.280 | 0.342 | 0.0834 | 0.0542 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: In the first two columns, the dependent variable is $\bar{\sigma}_z^{ann}$, the volatility of the zip-code-level annualized monthly log housing return. In the last two columns, the dependent variable is $\bar{r}_z^{ann}$, the average of that return. Returns are measured using the Zillow House Value Index for a cross-section of 5,573 zip codes from 1998-2015. The explanatory variable is the natural logarithm of the zip code's mean adjusted gross income (AGI) from 1998, as reported by the IRS. All regressions include MSA fixed effects. Standard errors are clustered by state. Columns (2) and (4) restrict to MSAs in which one million or more tax returns were filed with the IRS in 1998.

| Metro Area | Coefficient on Ln(Mean AGI) | Standard Error | Obs. | $R^2$ |
|---|---|---|---|---|
| New York | -0.0131*** | -0.00274 | 291 | 0.0735 |
| Los Angeles | -0.0102*** | -0.00147 | 289 | 0.143 |
| Chicago | -0.00828*** | -0.00235 | 213 | 0.0555 |
| Philadelphia | -0.0138*** | -0.00265 | 178 | 0.133 |
| Miami Fort Lauderdale | -0.00648*** | -0.00193 | 155 | 0.0689 |
| Atlanta | -0.0268*** | -0.00452 | 142 | 0.2 |
| Boston | -0.0153*** | -0.00287 | 147 | 0.165 |
| San Francisco | -0.0115*** | -0.00271 | 113 | 0.141 |
| Detroit | -0.0171*** | -0.00399 | 132 | 0.124 |
| Seattle | -0.00407 | -0.00248 | 112 | 0.0238 |
| Riverside | -0.0146*** | -0.0033 | 95 | 0.174 |
| Phoenix | -0.0176*** | -0.0034 | 89 | 0.236 |
| Minneapolis St Paul | -0.0127*** | -0.00348 | 104 | 0.116 |
| Tampa | -0.00675 | -0.00446 | 100 | 0.0228 |
| Baltimore | -0.0478*** | -0.0063 | 71 | 0.455 |
| Denver | -0.0155*** | -0.0046 | 81 | 0.125 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Each row corresponds to a cross-sectional regression of return volatility, based on CoreLogic data from 1998-2015, on 1998 mean household AGI, within one of the 16 largest MSAs in the sample.

| Metro Area | Coefficient on Ln(Mean AGI) | Standard Error | Obs. | $R^2$ |
|---|---|---|---|---|
| New York | -0.0240*** | -0.00208 | 291 | 0.315 |
| Los Angeles | -0.0341*** | -0.00255 | 289 | 0.384 |
| Chicago | -0.0192*** | -0.00274 | 213 | 0.19 |
| Philadelphia | -0.0196*** | -0.00222 | 178 | 0.306 |
| Miami Fort Lauderdale | -0.0321*** | -0.00284 | 155 | 0.455 |
| Atlanta | -0.0543*** | -0.0053 | 142 | 0.428 |
| Boston | -0.0242*** | -0.00228 | 147 | 0.437 |
| San Francisco | -0.0393*** | -0.00469 | 113 | 0.387 |
| Detroit | -0.0366*** | -0.00441 | 132 | 0.346 |
| Seattle | -0.00778*** | -0.0021 | 112 | 0.111 |
| Riverside | -0.0380*** | -0.00526 | 95 | 0.359 |
| Phoenix | -0.0375*** | -0.00435 | 89 | 0.462 |
| Minneapolis St Paul | -0.0313*** | -0.00497 | 104 | 0.28 |
| Tampa | -0.0257*** | -0.00409 | 100 | 0.288 |
| Baltimore | -0.0279*** | -0.00412 | 71 | 0.399 |
| Denver | -0.0250*** | -0.00352 | 81 | 0.39 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Each row corresponds to a cross-sectional regression of return volatility, based on Zillow data from 1998-2015, on 1998 mean household AGI, within one of the 16 largest MSAs in the sample.

| | (1) HPI return vol | (2) HPI return vol | (3) HPI return | (4) HPI return |
|---|---|---|---|---|
| Ln(Mean AGI) | -0.0130*** | -0.0149*** | 0.0000757 | -0.000993 |
| | (0.00159) | (0.00183) | (0.000826) | (0.00126) |
| Fixed effect | MSA | MSA | MSA | MSA |
| Obs. | 5438 | 2312 | 5438 | 2312 |
| $R^2$ | 0.166 | 0.212 | 0.0000112 | 0.00206 |

Standard errors in parentheses

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 5: In the first two columns, the dependent variable is the volatility of the zip-code-level *annual* log housing return. In the last two columns, the dependent variable is the average of that return. Returns are measured using the CoreLogic Home Price Index (Single Family Combined) for a cross-section of 5,573 zip codes from 1998-2015. The explanatory variable is the natural logarithm of the zip code's mean adjusted gross income (AGI) from 1998, as reported by the IRS. All regressions include MSA fixed effects. Standard errors are clustered by state. Columns (2) and (4) restrict to MSAs in which one million or more tax returns were filed with the IRS in 1998.

| | (1) ZHVI return vol | (2) ZHVI return vol | (3) ZHVI return | (4) ZHVI return |
|---|---|---|---|---|
| Ln(Mean AGI) | -0.0254*** | -0.0301*** | 0.00628*** | 0.00563*** |
| | (0.00378) | (0.00387) | (0.00122) | (0.00178) |
| Fixed effect | MSA | MSA | MSA | MSA |
| Obs. | 5438 | 2312 | 5438 | 2312 |
| $R^2$ | 0.320 | 0.393 | 0.0822 | 0.0594 |

Standard errors in parentheses

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 6: In the first two columns, the dependent variable is the volatility of the zip-code-level *annual* log housing return. In the last two columns, the dependent variable is the average of that return. Returns are measured using the Zillow House Value Index for a cross-section of 5,573 zip codes from 1998-2015. The explanatory variable is the natural logarithm of the zip code's mean adjusted gross income (AGI) from 1998, as reported by the IRS. All regressions include MSA fixed effects. Standard errors are clustered by state. Columns (2) and (4) restrict to MSAs in which one million or more tax returns were filed with the IRS in 1998.

|  | (1) HPI vol | (2) HPI vol | (3) HPI vol | (4) HPI vol |
| --- | --- | --- | --- | --- |
| $\Delta \ln(HPI)_{t-12}$ | -0.0384*** | -0.0357*** | -0.0357*** | -0.0362*** |
|  | (0.00721) | (0.00658) | (0.00688) | (0.00710) |
| $\Delta \ln(HPI)_{t-24}$ | -0.0225*** | -0.0216*** | -0.0221*** | -0.0223*** |
|  | (0.00548) | (0.00506) | (0.00504) | (0.00523) |
| $\Delta \ln(HPI)_{t-36}$ | -0.0108** | -0.00985* | -0.0100** | -0.0104* |
|  | (0.00530) | (0.00498) | (0.00494) | (0.00524) |
| $\Delta \ln(HPI)_{t-48}$ | -0.00905 | -0.00753 | -0.00838 | -0.00980 |
|  | (0.00847) | (0.00824) | (0.00845) | (0.00891) |
| Fixed effect | None | State | MSA | Zip Code |
| Obs. | 73828 | 73828 | 73427 | 73828 |
| $R^2$ | 0.0612 | 0.0855 | 0.129 | 0.0604 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(a) Using CoreLogic HPI

|  | (1) ZHVI vol | (2) ZHVI vol | (3) ZHVI vol | (4) ZHVI vol |
| --- | --- | --- | --- | --- |
| $\Delta \ln(ZHVI)_{t-12}$ | -0.0173* | -0.0215** | -0.0211** | -0.0206** |
|  | (0.00877) | (0.00972) | (0.00960) | (0.00944) |
| $\Delta \ln(ZHVI)_{t-24}$ | -0.0173*** | -0.0172*** | -0.0173*** | -0.0167*** |
|  | (0.00324) | (0.00330) | (0.00308) | (0.00311) |
| $\Delta \ln(ZHVI)_{t-36}$ | -0.00994** | -0.00864** | -0.00842** | -0.00803** |
|  | (0.00442) | (0.00370) | (0.00389) | (0.00365) |
| $\Delta \ln(ZHVI)_{t-48}$ | 0.0163 | 0.0127 | 0.0134 | 0.0129 |
|  | (0.0107) | (0.00908) | (0.00919) | (0.00947) |
| Fixed effect | None | State | MSA | Zip Code |
| Obs. | 73828 | 73828 | 73427 | 73828 |
| $R^2$ | 0.0135 | 0.152 | 0.224 | 0.0219 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(b) Using Zillow ZHVI

Table 7: Zip-code level panel regressions of housing return volatility (calculated on a rolling basis with 12 months of data) on lags of monthly housing returns.

| | (1) HPI vol | (2) HPI vol | (3) HPI vol | (4) HPI vol |
|---|---|---|---|---|
| $\Delta \ln(HPI)_{t-12}$ | -0.0361*** | -0.0338*** | -0.0344*** | -0.0352*** |
| | (0.00767) | (0.00714) | (0.00722) | (0.00736) |
| $\Delta \ln(HPI)_{t-24}$ | -0.0246*** | -0.0233*** | -0.0241*** | -0.0250*** |
| | (0.00657) | (0.00610) | (0.00607) | (0.00633) |
| $\Delta \ln(HPI)_{t-36}$ | -0.0155* | -0.0141* | -0.0149** | -0.0162** |
| | (0.00765) | (0.00731) | (0.00730) | (0.00764) |
| $\ln(AGI)_{1998}$ | -0.00618** | -0.00649*** | -0.00487*** | |
| | (0.00231) | (0.00230) | (0.00142) | |
| $\ln(AGI)_{1998} \times \Delta \ln(HPI)_{t-12}$ | 0.0161** | 0.0167** | 0.0157** | 0.0132** |
| | (0.00682) | (0.00680) | (0.00586) | (0.00530) |
| $\ln(AGI)_{1998} \times \Delta \ln(HPI)_{t-24}$ | 0.00100 | 0.00145 | 0.00101 | -0.000566 |
| | (0.00355) | (0.00339) | (0.00369) | (0.00367) |
| $\ln(AGI)_{1998} \times \Delta \ln(HPI)_{t-36}$ | -0.00743 | -0.00694 | -0.00807* | -0.00938* |
| | (0.00461) | (0.00440) | (0.00478) | (0.00544) |
| Fixed effect | None | State | Metro | Zip Code |
| Obs. | 76887 | 76887 | 76467 | 76887 |
| $R^2$ | 0.0634 | 0.0864 | 0.130 | 0.0631 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(a) Using CoreLogic HPI

| | (1) ZHVI vol | (2) ZHVI vol | (3) ZHVI vol | (4) ZHVI vol |
|---|---|---|---|---|
| $\Delta \ln(ZHVI)_{t-12}$ | -0.0122* | -0.0176** | -0.0175** | -0.0180** |
| | (0.00669) | (0.00749) | (0.00737) | (0.00727) |
| $\Delta \ln(ZHVI)_{t-24}$ | -0.0137*** | -0.0133*** | -0.0137*** | -0.0139*** |
| | (0.00306) | (0.00296) | (0.00296) | (0.00296) |
| $\Delta \ln(ZHVI)_{t-36}$ | -0.00124 | -0.00288 | -0.00239 | -0.00332 |
| | (0.00818) | (0.00647) | (0.00666) | (0.00668) |
| $\ln(AGI)_{1998}$ | -0.0115*** | -0.0107*** | -0.0106*** | |
| | (0.00140) | (0.00157) | (0.00154) | |
| $\ln(AGI)_{1998} \times \Delta \ln(ZHVI)_{t-12}$ | 0.0597*** | 0.0579*** | 0.0569*** | 0.0536*** |
| | (0.00782) | (0.00708) | (0.00728) | (0.00699) |
| $\ln(AGI)_{1998} \times \Delta \ln(ZHVI)_{t-24}$ | 0.0182*** | 0.0174*** | 0.0172*** | 0.0147*** |
| | (0.00573) | (0.00560) | (0.00559) | (0.00498) |
| $\ln(AGI)_{1998} \times \Delta \ln(ZHVI)_{t-36}$ | 0.00983 | 0.00802 | 0.00650 | 0.00425 |
| | (0.00835) | (0.00917) | (0.00865) | (0.00905) |
| Fixed effect | None | State | Metro | Zip Code |
| Obs. | 76887 | 76887 | 76467 | 76887 |
| $R^2$ | 0.0347 | 0.160 | 0.234 | 0.0347 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(b) Using Zillow ZHVI

Table 8: Repeats Table 7, interacting explanatory variables with 1998 log AGI.

| | (1) ZHVI vol | (2) ZHVI vol | (3) ZHVI vol | (4) LTV | (5) ZHVI vol |
|---|---|---|---|---|---|
| Non-recourse | 0.0335*** | 0.0318*** | 0.0317*** | -0.0221* | |
| | (0.00909) | (0.00866) | (0.00907) | (0.0109) | |
| LTV | | | | | -1.430* |
| | | | | | (0.795) |
| Housing expenditure share | 0.712** | 0.750** | 0.743* | -0.425** | 0.135 |
| | (0.303) | (0.283) | (0.380) | (0.182) | (0.470) |
| Ln(AGI) | | -0.0302*** | -0.0302*** | -0.0991*** | -0.172** |
| | | (0.00359) | (0.00312) | (0.00288) | (0.0790) |
| Ln(Population) | | | 0.000502 | -0.0145** | -0.0202 |
| | | | (0.00923) | (0.00537) | (0.0165) |
| Constant | -0.151 | -0.0474 | -0.0518 | 1.637*** | 2.289* |
| | (0.105) | (0.0942) | (0.0852) | (0.0841) | (1.298) |
| Sample | All | All | All | All | All |
| Obs. | 2891 | 2804 | 2804 | 2804 | 2804 |
| $R^2$ | 0.276 | 0.408 | 0.408 | 0.554 | . |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: This table reports zip-code-level regressions of return volatility on MSA-level average housing expenditure share and the state-level non-recourse indicator from Ghent and Kudlyak (2011). Standard errors clustered by state.

| | (1) Mean(Permits) | (2) Mean($\Delta$ log permits) | (3) $\sigma$(log permits) | (4) $\sigma$($\Delta$ log permits) |
|---|---|---|---|---|
| Ln(Mean AGI) | 1478.6$^{***}$ | -0.0171 | -0.0804$^{**}$ | -0.503$^{***}$ |
| | (164.3) | (0.0140) | (0.0341) | (0.0438) |
| Constant | -4917.6$^{***}$ | 0.0154 | 0.928$^{***}$ | 2.277$^{***}$ |
| | (544.8) | (0.0520) | (0.115) | (0.164) |
| Obs. | 2987 | 2924 | 2941 | 2902 |
| $R^2$ | 0.123 | 0.00130 | 0.00542 | 0.162 |

Standard errors in parentheses

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

(a)

| | (1) Mean(Permits) | (2) Mean($\Delta$ log permits) | (3) $\sigma$(log permits) | (4) $\sigma$($\Delta$ log permits) |
|---|---|---|---|---|
| Ln(Mean AGI) | 1646.8$^{***}$ | 0.00911 | -0.102$^{**}$ | -0.406$^{***}$ |
| | (211.9) | (0.0114) | (0.0431) | (0.0470) |
| Fixed effect | State | State | State | State |
| Obs. | 2987 | 2924 | 2941 | 2902 |
| $R^2$ | 0.128 | 0.000297 | 0.00750 | 0.0967 |

Standard errors in parentheses

$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

(b)

Table 10: Relationship between county-level income and permit issuance. In each regression the explanatory variable is the log of the county-level mean household AGI as reported in the IRS Statistics of Income for 1998. The outcome variables are county-level moments calculated from the county-level annual time series of permits issued for single-unit dwellings from 1998-2015: Average number of permits in column 1; average growth rate of log permits in column 1; volatility of log permits in column 3; volatility of the growth rate of log permits in column 4. All standard errors are clustered by state. Panel (b) includes state fixed effects in all four regressions.

| | (1) Age | (2) Post-2005 | (3) Age | (4) Post-2005 | (5) Age | (6) Post-2005 |
|---|---|---|---|---|---|---|
| Ln(HH Income) | -2.786*** | 0.0223*** | -2.330*** | 0.0200*** | | |
| | (0.233) | (0.00265) | (0.0257) | (0.000675) | | |
| Income bin 2 | | | | | -1.140*** | 0.00825*** |
| | | | | | (0.139) | (0.00167) |
| Income bin 3 | | | | | -2.753*** | 0.0217*** |
| | | | | | (0.254) | (0.00308) |
| Income bin 4 | | | | | -5.083*** | 0.0403*** |
| | | | | | (0.422) | (0.00502) |
| Income bin 5 | | | | | -8.059*** | 0.0634*** |
| | | | | | (0.613) | (0.00710) |
| Constant | 65.95*** | -0.156*** | 60.94*** | -0.131*** | 38.74*** | 0.0624*** |
| | (2.565) | (0.0291) | (0.282) | (0.00742) | (0.707) | (0.00341) |
| Fixed effect | State | State | 2000 PUMA | 2000 PUMA | | |
| Sample | 3525048 | 3525048 | 3525048 | 3525048 | 3525048 | 3525048 |
| Obs. | 0.0203 | 0.00522 | 0.0142 | 0.00426 | 0.0246 | 0.00643 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: This table reports property-level regressions of age (columns 1, 3, 5), and an indicator for being built after 2005 (columns 2, 4, 6) on measures of household income. Data are from the housing characteristics component of the American Community Survey (ACS) conducted by the US Census Bureau. In Columns 3 and 4, the fixed effect is at the level of Public Use Microdata Area as defined in the ACS. In Columns 5 and 6, bins of household income are constructed within-state. The consistent finding is that areas with higher income have newer housing stock.