Underdiversification and Housing Risk Premia

Esther Eiling   Erasmo Giambona   Ricardo Lopez A.   Patrick Tuijp

This Draft: October 17, 2023

Abstract

This paper studies the effect of underdiversification on risk premia in residential real estate, where constraints to diversification are relatively explicit but heterogeneous across investors. While households typically own one home, local landlords and national landlords hold more diversified housing portfolios. We first derive a simple theoretical model of multiple local housing markets with different types of investors facing different constraints to diversification. In equilibrium, zip code–level, local, and national housing risks are priced. The respective prices of risk depend, however, on the relative presence of underdiversified (homeowners) and diversified (local and national landlords) investors. We test the model using monthly zip code–level housing returns in 9,093 zip codes across 135 metropolitan statistical areas (MSAs), representing almost 70% of the U.S. population. We provides three main set of results. First, we document that in 71% of the MSAs at least one source of risk carries a significantly positive price of risk. Second, while zip code-level housing risk and local housing market risk are priced in about half of the MSAs, national housing market risk is only priced in about 19% of the MSAs. Third, in line with our model, we find that the propensity of zip code–level, local, and national housing risks to be priced is higher in MSAs where the presence of underdiversified (homeowners), local, and national landlords is higher, respectively.

Keywords: Underdiversification, Housing Risk Premia, Housing Market Segmentation, Homeowners, Local Landlords, National Landlords.

JEL Classification: G12, R30.

We thank Patrick Augustin, Shaun Bond, Steven Bourassa, Stefano Colonello, Joost Driessen, Marc Francke, Stuart Gabriel, David Geltner, Andra Ghent, Francesco Gomes, Lu Han, Gerard Hoberg, Martin Hoesli, Frank de Jong, Jack Liebersohn, Alberto Plazzi, Stephen Ross, Simon Rottke, Jacob Sagi, Sergei Sarkissian, Simon Stevenson, Johannes Stroebel, and conference and seminar participants at Binghamton University, Cornell University, ESSEC Business School, Fordham University, McGill University, Bonn University, Ortec Finance, Syracuse University, the University of Connecticut, the University of Torino, the 2019 ASSA-AREUEA and the 2019 AREUEA International Conference meetings for their helpful suggestions. We are very grateful to Aaron Franklin and Peter Gross at the Zillow Group for giving us access to their proprietary data. The results and opinions are those of the authors and do not reflect the position of Zillow Group. Eiling: University of Amsterdam, email: e.eiling@uva.nl; Giambona: Syracuse University, email: egiambon@syr.edu; Lopez: Syracuse University, email: rlopezal@syr.edu; Tuijp: Ortec Finance & University of Amsterdam, e–mail: patrick.tuijp@ortec–finance.com. Corresponding author: Esther Eiling, Finance Group, Amsterdam Business School, University of Amsterdam, Plantage Muidergracht 12, Postbus 15953, 1001 NL Amsterdam. Ph.: +31 020 525 4325. E–mail: e.eiling@uva.nl.

1 Introduction

Investor underdiversification is widespread. Studies have shown that retail investors (e.g., Blume and Friend, 1975; Kelly, 1995; Goetzmann and Kumar, 2008) hold underdiversified stock portfolios, also preferring domestic stocks ('home bias'), (e.g., French and Poterba, 1991; Tesar and Werner, 1998; Warnock, 2002).[1] Notably, portfolio underdiversification has also been documented for institutional investors (Koijen and Yogo, 2019). In this paper, we study the effect of underdiversification on asset pricing. While previous studies have focused on the impact of underdiversification on expected stock returns,[2] this study focuses on the effect of underdiversification on risk premia in residential real estate.

Housing offers a particularly suitable setting to study the link between underdiversification and risk premia for three main reasons. First, investing in residential real estate is characterized by explicit constraints to diversification. Housing is lumpy and most households who are owner–occupants only own one property (e.g., Flavin and Yamashita, 2002; Cocco, 2005; Giacoletti, 2021). At the same time, landlords and institutional investors can hold diversified housing portfolios within a certain region or nationally. Second, in housing, unlike the stock market, we can construct measures for the presence of underdiversified and diversified investors and test how such presence relates to the types of risk priced in the cross-section of housing returns. Third, we can contribute to the literature showing that households are exposed to significant idiosyncratic housing risk (e.g., Eichholtz et al., 2021; Giacoletti, 2021) by addressing an important, yet open question, on whether they are compensated for such risk, which has important welfare implications (e.g., Goetzmann and Kumar, 2003).

To help guide our empirical analysis, we start with a theoretical housing market model. The setup is stylized, focusing only on constraints to geographical diversification.[3] Our model contains

[1]While in principle stock diversification is straightforward (e.g., index ETF's), many investors choose to hold underdiversified stock portfolios, due, for instance, to skewness preferences (Mitton and Vorkink, 2007), behavioral biases, such as familiarity (Massa and Simonov, 2006) and probability weighting (Dimmock et al., 2020), costly information acquisition (Van Nieuwerburgh and Veldkamp, 2010), or crowding out effects from homeowneship (Flavin and Yamashita, 2002; Cocco, 2005).

[2]Papers such as Merton (1987), Della Preite et al. (2022), and Iachan et al. (2022) show that in the presence of underdiversification, asset–specific stock risk may be priced. Errunza and Losq (1985), De Jong and De Roon (2005), and Bekaert et al., 2009, among others, show that in the presence of cross-border diversification constraints local risk factors can be priced in international stock markets.

[3]To keep the model tractable, we leave out other aspects, such as leverage, transaction costs, short sales constraints and consumption good properties of housing. To the extent that residential real estate has investment good properties,

four local housing markets: two Metropolitan Statistical Areas (MSAs) with two zip codes each. There are three different types of investors who face different diversification constraints: households can only invest in one zip code (within one MSA), landlords can diversify across zip codes within one MSA, and institutional investors can diversify across all zip codes in both MSAs.[4] In this framework, local (MSA) housing markets are partially segmented. Housing risk premia are driven by three sources of risk: national housing market risk, MSA–specific housing market risk, and zip–code specific residual housing risk. The prices of risk depend on the relative presence of institutions, landlords, and households, who differ in their ability to hold a geographically diversified housing portfolio.

The resulting pricing equation is directly empirically testable using the standard two–pass regression approach of Fama and MacBeth (1973). We use monthly zip code–level housing returns from Zillow for 9,093 zip codes across 135 MSAs in the U.S.[5] We estimate the model for the cross–section of zip code housing returns within each local MSA separately. Hence, we estimate the model for 135 different cross–sections. Next, we link geographical variation in estimated prices of risk across these 135 MSAs to proxies for the presence of households, landlords, and institutions in each MSA. We use returns on national and MSA house value indices as national and local housing market risk factors, respectively, as well as zip code–level residual return volatility.

Our empirical analysis provides three main sets of results. First, we document substantial empirical support for our factor model in the cross–section of housing returns. In 96 of all 135 MSAs (71%), at least one source of risk carries a significant positive price of risk. This finding implies that for the majority of metropolitan areas, housing returns display investment–good properties, even at the zip code–level where many owner occupants also receive consumption benefits from homeownership.

Second, we find substantial variation across MSAs in terms of which sources of risk are priced. The two most important sources of priced risk are local MSA housing market risk and zip code–

we should see that, in the data, housing returns are captured by an asset pricing model.

[4]We consider zip codes as the most disaggregate housing investment, rather than individual properties, for two reasons. First, this matches our empirical analysis, which is based on monthly zip code level housing returns. Ideally, we would consider property level returns, but due to infrequent trading, monthly housing returns are not available. This makes performing asset pricing tests and estimating risk exposures and expected returns infeasible. Second, modeling zip code–level housing returns avoids the issue of indivisibility of houses.

[5]Our data runs from April 1996 to December 2016 and represents almost 70% of the U.S. population. Several papers have used zip code–level or other aggregation–level data from Zillow to analyze mortgage defaults, foreclosures, and household leverage (e.g., Mian and Sufi, 2009; Mian et al., 2015; Adelino et al., 2016).

specific housing residual risk. They are priced in 49 and 48 MSAs, respectively, or about half of the 96 MSAs where housing has investment good properties. In contrast, despite the finding that zip code housing returns have relatively high and dispersed exposures to U.S. housing market risk (the average beta is 1.05 and the cross–zip code dispersion is 0.69), national U.S. housing market risk plays a much smaller role in the cross–section of housing returns. It is priced only in 19% of the MSAs in our sample. Hence, even though aggregate U.S. house price indexes are often used to assess housing market movements, for most owner–occupiers, national housing risk is not priced.

Heatmaps reveal substantial geographical variation in terms of which sources of risk are priced in each local housing market. Our third set of results links this variation to the presence of investors who differ in their ability to hold a diversified housing portfolio. We proxy for the presence of households using homeownership data from the U.S. Census and American Community Survey. We proxy for the presence of local landlords using data from a range of sources, including the U.S. Census Bureau and the S&P Global Market Intelligence SNL Real Estate Property database. Finally, we proxy for the presence of national institutional investors using residential real estate holdings data of real estate companies from the S&P Global Market Intelligence SNL Real Estate Property database.

Our model predicts that in MSAs where there are more underdiversified investors, zip code–level housing risk carries a higher price of risk. We test this prediction by regressing in a probit setting an indicator for whether zip code–level risk is significantly priced in an MSA on an indicator for high homeownership, our proxy for the presence of underdiversified investors, and control variables. We find that zip code–level housing risk is 29.4% more likely priced in MSAs where homeownership is high. We also find that in MSAs where there are more local landlords, local MSA–level housing risk is significantly more likely to be priced. In line with our model, we further find that national housing risk is more likely priced in MSAs with significant presence of institutional investors (national landlords).[6] Overall, with the caveat that caution should be used in drawing causality conclusions, our results indicate that the types of risk priced across local housing markets are related to the types of investors active in those markets.

Our findings pass a battery of robustness tests. First, we find that our results hold when we

[6]While several studies have shown that the presence of institutional investors or out–of–town investors drives up house prices (e.g., Chinco and Mayer, 2016; Badarinza and Ramadorai, 2018; Mills et al., 2019; Favilukis and Van Nieuwerburgh, 2021), they have not focused on how such presence affects housing risk premia.

estimate our own zip code–level house price indexes. To build these indexes, we use proprietary property-level data including home characteristics from over 400 million detailed public records from the Zillow Transaction and Assessment Dataset (ZTRAX). Second, our results are robust to the inclusion of stock market returns in our multi–factor model. In line with Jordà et al. (2019), we find that, at the zip code level, housing returns are not exposed to stock market returns, with betas very close to zero. Further, our main conclusions hold under the following additional conditions: 1) when we consider two sub sample periods, before and after the sub prime mortgage crisis; 2) when we increase the number of MSAs by selecting those with a minimum of 15 zip codes instead of 20; 3) when we control for housing value and zip code size; and 4) when we unsmooth housing returns. We discuss these robustness tests and many other robustness tests later in the paper and in detail in the Online Appendix.

Our paper relates to the literature on underdiversification and asset pricing. This literature has shown theoretically (e.g., Levy, 1978; Merton, 1987; Malkiel and Xu, 2004; Della Preite et al., 2022; Iachan et al., 2022) and empirically (e.g., Ang et al. (2006), Huang et al. (2010), Stambaugh et al. (2015), and Herskovic et al. (2016)) that if investors face diversification constraints stock idiosyncratic risk will priced. Relatedly, the international finance literature has further shown that barriers to international investments lead to partially or fully segmented markets in which local risk is priced (e.g., Errunza and Losq, 1985; De Jong and De Roon, 2005; and Bekaert et al., 2009). We contribute to this literature by focusing on the housing market, a setting that, unlike the stock market, allows us to assess the presence of underdiversified and diversified investors and relate this presence to the types of risk priced in the cross-section of housing returns.

Our paper is also related to the literate on the risk–return trade off in real estate returns. Papers that analyze the risk–return trade–off in real estate tend to focus either on commercial real estate (e.g., Plazzi, Torous and Valkanov, 2008), or rely on aggregate housing returns at the national or metropolitan levels (e.g., Case, Cotter and Gabriel, 2011; Han, 2013; Cotter, Gabriel and Roll, 2015). These aggregate housing returns are essentially based on a broadly diversified housing portfolio within an MSA that can be as large as the combined area of New York City, Newark, and Jersey City. However, individual homeowners do not experience these diversified MSA–level housing returns. In fact, we document substantial variation in average housing returns and risk exposures across zip codes within a given MSA.[7] Our results show that in the presence of homeowner underdiversification and housing market segmentation, the risk and return characteristics of disaggregate housing returns differ from those of the aggregate housing market.

Our paper is more closely related to the literature on idiosyncratic housing risk. Giacoletti (2021) and Eichholtz et al. (2021) find that property-level idiosyncratic housing risk constitutes the largest share of total housing risk. Sagi (2021) finds similar results for commercial real estate.[8] We focus on zip code–level data and in line with these papers we find that on average, zip code–specific volatility accounts for the majority (63%) of total volatility. More importantly, we complement this literature by showing that underdiversified homeowners are compensated for zip code specific risk.[9] To our best knowledge, our paper is the first to document empirically this important finding.

The remainder of the paper is organized as follows. Section 2 derives a simple theoretical asset pricing model with underdiversification that serves as a basis for our empirical analysis. Section 3 discusses the data and descriptive statistics. In Section 4, we discuss the main empirical results. Robustness tests are in Section 5. Section 6 concludes. The Appendix provides additional details of the model derivations and the data, as well as additional results. An Online Appendix reports detailed robustness tests.

2 A Stylized Asset Pricing Model with Underdiversification

As a basis for our empirical analysis, we first derive a stylized asset pricing model for the cross–section of expected housing returns. To keep the model tractable, we focus on only one particular aspect of the housing market, underdiversification. This results in a pricing equation that is directly empirically testable. The model also provides testable implications on the link between the presence of underdiversified investors and the types of risk priced.

[7] These findings are in line with the literature showing that housing markets cluster and local aspects matter. This literature includes Goetzmann, Spiegel and Wachter (1998); Genesove and Han (2007); Han and Strange (2014); Landvoigt, Piazzesi and Schneider (2015); Tuzel and Zhang (2017); Piazzesi, Schneider and Stroebel (2020), Van Nieuwerburgh and Weill (2010); and Ghent (2021).

[8] Landvoigt et al. (2015) analyze the cross–section of house capital gains within the San Diego MSA and find that access to cheap credit for poor households was a major driver of capital gains in the low end of the market. Several recent papers analyze the effect of climate change on house prices, such as Bernstein et al. (2019), Murfin and Spiegel (2020), Baldauf et al. (2020) and Giglio et al. (2021).

[9] Cannon et al. (2006) also consider zip code–level housing returns but have a limited time series dimension of eight annual observations, which complicates asset pricing tests. More broadly, our paper is related to the studies on the determinants of idiosyncratic volatility of house prices (Peng and Thibodeau, 2017) and stock market betas of property–level housing returns (Peng and Zhang, 2021).

The model has three key features. First, we consider housing as a risky asset. In other words, we focus on the investment good properties of housing and abstract away from consumption good properties.[10] If housing were merely a consumption good, a risk–based asset pricing model would not capture cross–sectional differences in expected housing returns. Second, we allow for the presence of different types of investors who face different constraints to geographical diversification of their housing investments. Third, we focus on zip codes instead of individual properties as the most disaggregate local housing returns. This allows us to avoid having to deal with indivisibility of individual properties in the mode and matches our data, which is at the zip code level.

Our model builds upon Malkiel and Xu (2004). In a mean–variance framework, these authors show that stock idiosyncratic risk is priced in the presence of underdiversified investors (i.e., investors who cannot hold diversified stock portfolios). In our setting, we add a local market layer and link the type of restrictions investors face to their status as homeowners, local landlords, or institutional investors (national landlords). Our framework also relates to international asset pricing models where only a subset of stocks are tradeable by local investors compared with gloabl investors (e.g., De Jong and De Roon, 2005). The key difference is that the diversification constraints in our setting are not at the asset level, but at the investor level.

2.1 Model Setup

Consider a national housing market with four risky assets divided over two local metropolitan areas, i.e., MSAs. MSA A and MSA B each have two zip codes 1 and 2. Define expected housing returns $\mu = E[R]$ at the zip code level and the covariance matrix Σ as

$$\mu = \begin{bmatrix} \mu_{A1} \\ \mu_{A2} \\ \mu_{B1} \\ \mu_{B2} \end{bmatrix} = \begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix} \qquad \text{and} \qquad \Sigma = \begin{bmatrix} \Sigma_A & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_B \end{bmatrix}, \tag{1}$$

[10] As discussed in the introduction, we also abstract away from other aspects of the housing market, such as leverage, short sales constraints, liquidity issues, etc. This allows us to keep the model tractable, derive a pricing equation that is empirically testable and focus purely on the asset pricing implications of the presence of underdiversified investors.

where

$$\Sigma_A = \begin{bmatrix} \sigma^2_{A1} & \sigma_{A1A2} \\ \sigma_{A2A1} & \sigma^2_{A2} \end{bmatrix} \quad \text{and} \quad \Sigma_{AB} = \begin{bmatrix} \sigma_{A1B1} & \sigma_{A1B2} \\ \sigma_{A2B1} & \sigma_{A2B2} \end{bmatrix}. \tag{2}$$

We use similar definitions for Σ_{BA} and Σ_B. We denote the covariance between housing returns in zipcode A1 and B1 as σ_{A1B1} and maintain similar definitions for the other covariance terms. We also include a risk-free bond with return r.

There are three types of investors who face different constraints to geographical diversification in the housing market. First, households are constrained to invest only in a single zip code within an MSA. We denote the total number of households in each of the four zip codes by $N_{H_{A1}}$, $N_{H_{A2}}$, $N_{H_{B1}}$, and $N_{H_{B2}}$. The second type of investors are local landlords, who are restricted to invest in one MSA, but can diversify across zip codes within the MSA. The number of landlords in each MSA is denoted by N_{L_A} and N_{L_B}. Finally, our setup includes N_I institutional investors (national landlords) who are unconstrained and can invest in all zip codes across both MSAs. It will be useful to write $N_{H_A} = N_{H_{A1}} + N_{H_{A2}}$ for the total number of households in MSA A (similarly for N_{H_B}), and $N_C = N - N_I$ for the total number of constrained investors (i.e., local landlords and households in both MSAs), where N denotes the total number of investors. We assume that all investors have mean-variance utility with risk aversion $1/\tau$.

For ease of notation, we introduce two matrices to select elements corresponding to each MSA, which are given by

$$G_A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad G_B = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

such that $\Sigma_A = G_A \Sigma G'_A$ and $\mu_A = G_A \mu$ and similarly, $\Sigma_B = G_B \Sigma G'_B$ and $\mu_B = G_B \mu$.

2.2 National Market Clearing

The investors form mean-variance optimal portfolios, given their spatial constraints, with portfolio weights given by

$$
\begin{aligned}
X_{H_{A1}} &= \tau \begin{bmatrix} \sigma_{A1}^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} (\mu - r\iota) \\
X_{L_A} &= \tau \begin{bmatrix} \Sigma_A^{-1} & 0 \\ 0 & 0 \end{bmatrix} (\mu - r\iota) \\
X_I &= \tau \Sigma^{-1} (\mu - r\iota),
\end{aligned} \tag{4}
$$

where ι is a 4×1 vector of ones. Similar expressions hold for households and local landlords in the other zip codes and in MSA B.

National aggregate demand D is given by the sum of all investor demand, weighed by the number of each type of investors. The vector D can be expressed as

$$D = \tau N \left\{ \frac{N_C}{N} \Sigma^{*-1} + \frac{N_I}{N} \Sigma^{-1} \right\} (\mu - r\iota),$$

where Σ^{*-1} represents a weighted sum of the relevant inverse covariance matrices these different groups of constrained investors face:

$$\Sigma^{*-1} = \left[\frac{N_{H_A}}{N_C} G_A' \Sigma_{H_A}^{-1} G_A + \frac{N_{H_B}}{N_C} G_B' \Sigma_{H_B}^{-1} G_B + \frac{N_{L_A}}{N_C} G_A' \Sigma_A^{-1} G_A + \frac{N_{L_B}}{N_C} G_B' \Sigma_B^{-1} G_B \right]$$

and

$$\Sigma_{H_A}^{-1} = \begin{bmatrix} \frac{N_{H_{A1}}}{N_{H_A}} (\sigma_{A1}^2)^{-1} & 0 \\ 0 & \frac{N_{H_{A2}}}{N_{H_A}} (\sigma_{A2}^2)^{-1} \end{bmatrix}.$$

Denote S the national aggregate supply vector of all four assets. Imposing aggregate market clearing, such that $D = S$, leads to the following pricing equation

$$
\begin{aligned}
(\mu - r\iota) &= \frac{1}{\tau N}\left[\Sigma^{-1} - \frac{N_C}{N}(\Sigma^{-1} - \Sigma^{*-1})\right]^{-1} S \\
&= \frac{1}{\tau N}\Sigma S + \frac{N_C}{N^2\tau}\Sigma\omega,
\end{aligned} \tag{5}
$$

where

$$
\omega = \left[(I - \Sigma^{*-1}\Sigma)^{-1} - \frac{N_C}{N} I\right]^{-1} S,
$$

which, following Malkiel and Xu (2004), can be interpreted as adjusted supply. The second term of the pricing equation naturally becomes zero in absence of constrained investors.

Next, denote α as the vector of weights in the total housing market portfolio: $\alpha = S/M$ with $M = \iota' S$ as the total market capitalization. The expected return on the market portfolio equals $\mu_m = \alpha'\mu$ and its variance $\sigma_m^2 = \alpha'\Sigma\alpha$. The betas with respect to the total market portfolio are defined as usual as

$$
\beta = \frac{1}{\sigma_m^2}\Sigma\alpha.
$$

With market clearing at the aggregate national level, we can express the risk premia of housing returns at the zip code level as a function of their exposures to national housing market risk, β, and residual variance, measured with respect to the national market portfolio. The following proposition, for which the proof is given in Appendix A.1, formalizes this first result.

Proposition 2.1. *Imposing national market clearing, such that $D = S$, leads to the following pricing equation*

$$
\begin{aligned}
(\mu - r\iota) &= \beta(\mu_m - r) + \frac{(\mu_m - r)/\sigma_m^2}{\left(1 + \frac{N_C}{N}\beta'\omega_*\right)}\frac{N_C}{N}\left(\Sigma - \beta\sigma_m^2\beta\right)\omega_* \\
&= \beta(\mu_m - r) + \chi\sigma_{SR}\Sigma_\epsilon^{nat}\omega_*,
\end{aligned} \tag{6}
$$

where $\Sigma_{\epsilon}^{nat} = \left(\Sigma - \beta\sigma_m^2\beta'\right)$ *and* $\omega_* = \frac{1}{M}\omega$. *The constant terms are defined as*

$$\chi = \frac{N_C}{N\left(1 + \frac{N_C}{N}\beta'\omega_*\right)}$$

$$\sigma_{SR} = (\mu_m - r)/\sigma_m^2.$$

This expression shows that in the presence of constrained investors who can only invest in a subset of the assets, i.e., local landlords and/or households, we have that $N_C/N > 0$ and residual risk beyond national market risk (i.e. Σ_{ϵ}^{nat}) is priced. The above pricing equation is similar to Malkiel and Xu (2004), except that the relative supply adjustment ω_* and the structure of Σ_{ϵ}^{nat} are different. The reason is that our model has an extra layer because we have two types of diversification constraints, at the MSA level and at the zip code level. Therefore, in a next step, we need to impose market clearing at the local MSA level. As we show in the next sub section, this leads to partially segmented local housing markets in which national market risk, local MSA–level market risk, and zip code–specific idiosyncratic housing risk are priced.

2.3 Local Market Clearing

Before imposing local market clearing at the MSA level, we further decompose risk beyond national market risk, Σ_{ϵ}^{nat}, into MSA–level market risk and zip code–level residual risk. To this end, we assume that Σ_{ϵ}^{nat} is block diagonal and that all covariances between housing returns across MSAs are captured by their exposures to common national market risk:

$$\Sigma_{\epsilon}^{nat} = \begin{bmatrix} \Sigma_{\epsilon A}^{nat} & O \\ O & \Sigma_{\epsilon B}^{nat} \end{bmatrix}.$$

In what follows, we will focus on MSA A, yet by symmetry of our setup, similar expressions hold for MSA B. The weights in the local MSA value–weighted market portfolio returns are given by $\alpha_{mA} = S_A/M_A$, where $S_A = G_A S$ and M_A is the total market capitalization in MSA A. As such we have $\mu_{mA} = \alpha_{mA}\mu_A$ and $\sigma_{mA}^2 = \alpha_{mA}'\Sigma_A\alpha_{mA}$. The exposures of zip code–level returns in MSA

A with respect to local market returns are defined as usual as

$$\beta_{mA} = \frac{1}{\sigma^2_{mA}} \Sigma_A \alpha_{mA}.$$

To distinguish between the asset pricing effects of national market risk and local MSA market risk, we need to account for the correlation between MSA market and national market returns.[11] We therefore decompose local market return variance σ^2_{mA} into the variance due to the exposure to the national market return and the so–called orthogonalized local market return variance, $\sigma^2_{mA\_orth}$, as follows:

$$\sigma^2_{mA} = \beta^2_{mA,m} \sigma^2_m + \sigma^2_{mA\_orth}, \tag{7}$$

where $\beta_{mA,m}$ is the exposure of the local market portfolio return with respect to the national market portfolio return, i.e., $\beta_{mA,m} = cov(R_{mA}, R_m)/\sigma^2_m$.

Next, we adjust the local market betas β_{mA} for the exposure of the local market returns with respect to national market returns, leading to so–called orthogonalized local market betas:

$$\beta^{orth}_{mA} = \beta_{mA} \frac{\sigma^2_{mA}}{\sigma^2_{mA\_orth}} - \beta_{mA,m} G_A \beta \frac{\sigma^2_m}{\sigma^2_{mA\_orth}}. \tag{8}$$

We can now decompose residual variance of housing returns in MSA A with respect to the national market portfolio, $\Sigma^{nat}_{\epsilon A}$, into an MSA–specific component and zip code–level residual risk:

$$\Sigma^{nat}_{\epsilon A} = \beta^{orth}_{mA} \sigma^2_{mA\_orth} \beta^{orth\prime}_{mA} + \Sigma_{\epsilon A} \tag{9}$$

where $\Sigma_{\epsilon A}$ is a 2×2 matrix of zip code–level residual variance of the zip codes in MSA A, which is idiosyncratic to national and MSA–level market risk.

By imposing local market clearing within MSA A, we can write the expected excess housing returns for all zip codes in MSA A as a function of i) their exposures to national market risk β, ii) exposures to local MSA–level market risk (orthogonalized with respect to the national market returns) β^{orth}_{mA}, and iii) zip code–level idiosyncratic housing risk $\Sigma_{\epsilon A}$. The following proposition, for which the proof is given in Appendix A.1, formalizes this result.

[11] See also, e.g., Bekaert, Hodrick and Zhang (2009); Conrad, Dittmar and Ghysels (2013).

Proposition 2.2. *Imposing local MSA market clearing, $D_A = S_A$, results in the following pricing equation for expected housing returns in MSA A:*

$$\begin{aligned}
\mu_A - r\iota = {} & G_A \beta \left(\mu_m - r\right) \\
& + \frac{N_C}{N} \left(\beta_{mA}^{orth\prime} G_A \omega_*\right) \frac{\sigma^2_{mA_orth}}{\sigma^2_{mA,adj}} \frac{M}{M_A} \frac{N_A}{N} \beta_{mA}^{orth} \left(\mu_{mA} - r\right) \\
& + \frac{N_C}{N} \sigma_{SR,A_adj} \frac{M}{M_A} \frac{N_A}{N} \Sigma_{\varepsilon A} G_A \omega_*,
\end{aligned} \tag{10}$$

where

$$\begin{aligned}
\sigma_{SR,A_adj} &= \frac{\mu_{mA} - r}{\sigma^2_{mA,adj}}. \\
\sigma^2_{mA,adj} &= \alpha_{mA}' \Sigma_A^* S_{adjA} \frac{1}{M_A} \\
S_{adjA} &= S_A - \tau N_I G_A \Sigma^{-1} (\mu - r\iota) \\
\Sigma_A^{*-1} &= \Sigma_A^{-1} - \frac{N_{H_A}}{N_A} \left(\Sigma_A^{-1} - \Sigma_{H_A}^{-1}\right)
\end{aligned}$$

To provide more economic intuition, note that S_{adjA} captures the total supply of housing in MSA A adjusted for the demand for those assets from institutional investors who can also invest in the other MSA. With local MSA housing market clearing, S_{adjA} equals the demand for housing in MSA A from all investors who are constrained to invest only in MSA A, the households and the local landlords in MSA A. Σ_A^{*-1} can be expressed as the weighted sum of inverse covariance matrices these constrained investors in MSA A face.

2.4 Housing Risk Premiums and the Presence of Underdiversified Investors

To make the above pricing equation directly empirically testable, we assume that the zip code residual covariance matrix $\Sigma_{\varepsilon A}$ is diagonal. In other words, we assume that all commonality between zip code housing returns is captured by the national and MSA–level market factors. Following Malkiel and Xu (2004), we assume that $\omega_{*,i}$ and $\sigma^2_{\varepsilon i}$ are independent in the cross–section, such that we can estimate the pricing equation using a standard two-pass cross–sectional regression of the form

$$E[R_{Ai} - r] = \lambda_0 + \lambda_{US} \beta_{Ai} + \lambda_{MSA\_A} \beta^{orth}_{mA,Ai} + \lambda_{ZIP\_A} \sigma^2_{\varepsilon Ai}, \tag{11}$$

where $E[R_{Ai} - r]$ denote expected excess returns for zip code i in MSA A. λ_0 should equal zero and the prices of risk equal

$$\lambda_{US} = (\mu_m - r) \tag{12}$$

$$\lambda_{MSA\_A} = \frac{N_C}{N}\left(\beta_{mA}^{orth\prime} G_A \omega_*\right) \frac{\sigma^2_{mA\_orth}}{\sigma^2_{mA,adj}} \frac{M}{M_A} \frac{N_A}{N} (\mu_{mA} - r) \tag{13}$$

$$\lambda_{ZIP\_A} = \frac{N_C}{N} \sigma_{SR,A\_adj} \frac{M}{M_A} \frac{N_A}{N} \bar{\omega}_{A*}, \tag{14}$$

where we denote $\bar{\omega}_{A*}$ as the average of all elements in $G_A \omega_*$:

$$\bar{\omega}_{A*} = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} G_A \omega_*. \tag{15}$$

A similar expression holds for excess zip code level returns in MSA B.

In our empirical analysis, we will estimate the linear factor model in eq. (11) for the cross–section of zip code–level housing returns within each MSA separately. The model also gives us testable implications for the link between the relative importance of three types of risk and the presence of different types of investors in each MSA.

First, if there are no constrained investors (i.e., in the presence of only institutional investors), then we are back at the standard CAPM at the national level. On the other hand, if there are no institutional investors, there is only local MSA–level market clearing and national housing market risk λ_{US} is not priced.

As we show in Appendix A.3, the prices of local MSA– and zip code–level risk depend on the relative presence of constrained investors. When there are more constrained investors in MSA A who cannot fully diversify[12] across all local MSAs and zip codes (i.e., N_A/N increases), the corresponding prices of local risk in that MSA also increase. In other words, if there are relatively more institutional investors who can achieve maximum diversification across all local markets (i.e., N_C/N decreases), the price of local risk will decrease. In the Appendix we also show that the presence of households matters more for the price of zip code–level residual risk, while the presence of local landlords matters relatively more for the price of local MSA–level housing risk.

[12]Given the limited number of assets available in our setup even to the institutional investors, in our context we mean by full diversification that the impact of the zip code–level risk on the portfolio variance is reduced as much as possible.

In our empirical tests, we link the cross–MSA variation in estimated prices of national, local MSA, and residual zip code–level housing risk to proxies for the presence of institutional investors (national landlords), local landlords, and households, respectively.

2.5 Empirical Implementation

We can test the resulting multi–factor asset pricing model using standard two–pass regressions of Fama and MacBeth (1973). In our main empirical specification, we do not control for housing returns' potential exposure to the equity market by including a U.S. equity market risk factor.[13] Due to partial local market segmentation, we have different pricing equations for different MSAs. We therefore separately estimate the model for the cross–section of zip codes within each MSA in our sample.

In the first step, we estimate the factor exposures and zip code–level residual volatility by the following time series regression:

$$r_{i,j,t} = \alpha_{i,j} + \beta_{i,j}^{US} r_t^{US} + \beta_{i,j}^{MSA} r_{MSA\_j,t}^{orth} + \varepsilon_{i,j,t}, \tag{16}$$

where $r_{i,j,t}$ denotes the housing return of zip code i in MSA j in month t in excess of the risk–free rate, r_t^{US} is the U.S. housing market excess returns, and $r_{MSA\_j,t}^{orth}$ is the orthogonalized excess housing market return for MSA j. In line with our theoretical model, these orthogonalized local MSA housing market returns are estimated using the following ordinary least squares regression:

$$r_{MSA\_j,t} = \mu_j + \beta_j r_t^{US} + \eta_{j,t}, \tag{17}$$

where the error term $\eta_{j,t}$ is our measure for orthogonalized local market returns $r_{MSA\_j,t}^{orth}$. Given that our full sample contains 20 years of monthly returns, as we discuss in the next section, we estimate Eq. (16) using the full sample period. We measure zip code–level residual risk as the standard deviation of the residuals: $\sigma_{i,j}^{ZIP} = stdev(\varepsilon_{i,j,t})$.

In the second step, we estimate the MSA–specific prices of risk for each factor by the following

[13] In a robustness test, we include the equity market factor and find similar results.

cross–sectional regression for a given MSA j:

$$E[r_{i,j}] = \lambda_j^0 + \lambda_j^{US}\beta_{i,j}^{US} + \lambda_j^{MSA}\beta_{i,j}^{MSA} + \lambda_j^{ZIP}\sigma_{i,j}^{ZIP}, \tag{18}$$

where we only include zip codes i that belong to MSA j and we use the risk exposures and zip code–level residual volatilities as estimated using Eq. (16).

In a final step, we test the link between the relative presence of households, local landlords and institutional investors (national landlord) in each MSA to the estimated prices of risk in the MSA. More specifically, we estimate a probit regression in which the dependent variable is an indicator equal to one if the price of risk estimate (for a given source of risk in a given MSA) is positive (as indicated by our model) and significant, and zero otherwise. The independent variables include proxies for the presence of different groups of investors, which we discuss in the next section.

3 Data and Descriptive Statistics

In this section, we discuss our data sources and basic descriptive statistics. Detailed variable definitions are in Table A.1.

3.1 Housing Return Data

Housing returns are from Zillow.[14] Zillow provides monthly Zillow Home Value Index (ZHVI) for different aggregation levels, ranging from zip codes to the states. We use the ZHVI for all homes, which includes single family residences, condominiums, and coops. The ZHVI is based on estimates on the market value of individual homes that Zillow calls Zestimates, which are based on a hedonic model that includes both appraisal and transaction data on house prices.[15] In line with our theoretical model, we focus on zip code–level housing returns, which is the lowest aggregation level of ZHVI available. We merge the zip code–level data with the MSA–level ZHVI data. We are left with 12,243 unique zip codes and 571 unique MSAs for the sample period from April 1996 to

[14]Data is downloaded on April 4th, 2017.

[15]Zillow estimates Zestimates using a proprietary model, but they do not disclose the model specification. Therefore, in Section 5.1, we build our own zip code–level house price indexes using property–level data from the Zillow Transaction and Assessment Dataset (ZTRAX). All our main results are robust to the use of these alternative house price indexes.

December 2016.[16] Zillow also provides the ZHVI for the entire U.S. residential real estate market, which we use in our empirical analysis as a proxy for aggregate U.S. housing market prices. To compute housing excess returns, we subtract the one–month Treasury bill rate from the monthly housing returns.[17] These data are obtained from the online data library of Kenneth R. French.[18]

The zip code–level data from Zillow do not cover all zip codes in each MSA. To assess of how well the Zillow zip codes cover each MSA, we obtain population data per zip code from the 2010 U.S. Census. We merge these data with the Zillow zip code–level data and compute the population per MSA. We then compare this population in relation to the reported MSA total population from the 2010 U.S. Census. We find that, on average, the Zillow zip code–level data covers 86.5% of each MSA's total population, with a median coverage of 91.6%. The Zillow zip code data thus provides good coverage for each MSA.

Since our multifactor model specified in Eq. (18) includes three systematic risk factors and zip code-level volatility, we need a sufficiently large cross section of zip codes to estimate the prices of risk. Thus, we keep only MSAs that have at least 20 zip codes, which reduces our sample size to 9,093 zip codes across 135 MSAs. However, these 135 MSAs represent 86.9% of the total population of the initial sample of 571 MSAs, or 67.9% of the entire U.S. population.

3.2 Homeownership and Rental Data

Our model predicts that the pricing of different sources of housing risk is related to the relative presence of housing investors with different degree of underdiversification: (1) owner–occupier households, (2) local landlords, and (3) institutional investors (national landlords). We therefore need ownership data for residential real estate at the household, local landlord, and national landlord levels.

To measure homeownership by households at MSA level, we use data from the U.S. Census and the American Community Survey (ACS) obtained through the Integrated Public Use Microdata Series (IPUMS) database (see, Ruggles et al. 2017). Our IPUMS dataset consists of annual data

[16] In a robustness test (Section 5), we consider the sample period ending in 2007 to exclude the subprime financial crisis.

[17] The total housing return should include rental yields as well. However, data on rents at the zip code-level is not currently available. Eichholtz et al. (2021) show that ignoring rental yields leads to underestimating idiosyncratic risk. Thus, our estimates of zip code-level housing residual risk and the pricing of this risk can be considered conservative due to the lack of data on rental yields.

[18] See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

for 290 MSAs from 2000 (decennial U.S. Census 1% sample) and 2005–2016 (ACS 1% samples). To identify MSAs in IPUMS, we use the 2013 definitions from the U.S. Office of Management and Budget (IPUMS item MET2013 which is available only for the 2000, and 2005–2016 samples). MSA homeownership is computed as the percentage of owner-occupiers (IPUMS item OWNERSHP) relative to all households per MSA. We obtain a yearly time–series of homeownership estimates for the sample of 2000, and 2005–2016. Finally, MSA homeownership is the average yearly homeownership.[19] Using the homeownership variable, we build "High–Homeownership", a dummy variable equal to one for MSAs with an homeownership rate above the sample top quartile, and zero otherwise. In line with our model, because households typically own one home, high homeownership indicates a higher degree of household underdiversification within an MSA and hence a higher propensity for zip-code specific residual housing risk to be priced.

To measure the relative presence of local landlords (those investing in multiple zip codes within an MSA), we proceed as follows. For each MSA, we first multiply the total number of housing units (U.S. Census Bureau) by one minus the MSA homeownership rate. From this number, we subtract the MSA stock of second home units (IPUMS NHGIS database, Manson et al., 2022). We do so because second homes are held by households and therefore are not part of the total number of rental residential (housing or apartment) units. Finally, we subtract the total number of residential units held by national landlords (real estate companies holding residential units in multiple MSAs), measured as discussed in detail below. This gives us the total number of rental residential units held by local landlords, which we divide by the total number of residential units in the MSA: local-landlord residential ownership. Using this ratio, we build "Local–Landlord", an indicator for MSAs with local-landlord residential ownership above the sample top quartile. As predicted by our model, we expect the MSA housing risk to be more likely priced in MSAs with a significant presence of local landlords.

Finally, to proxy for the presence of institutional investors (national landlords), those investing in multiple MSAs, we gather data on all public and private real estate companies, including Real Estate Investment Trusts (REITs) and Real Estate Operating Companies (ROCs), from the S&P Global Market Intelligence SNL Real Estate Property database over the period 1996–2016. The

[19]Since we estimate unconditional prices of risk in our asset pricing tests, we can only link cross–MSA variation in these prices of risk to ownership proxies. Therefore, we take time series averages of these proxies at the MSA level.

database contains the most recent information on the properties held by the real estate companies, including property type (e.g., residential, office, etc.), address, acquisition date, sale date, etc. For our purposes, we only focus on real estate companies holding residential properties. Using property transaction dates, we are able to build a time-series of residential property portfolios for all the real estate companies in our sample during 1996–2016.

There are 77 real estate companies holding residential properties in the SNL database, of which 32 are public and 45 private. To give some perspective on the relative size of national landlords in the U.S. residential market, we note that in 2016 (the last year in our sample), these entities hold 2.7 million housing units, which is about 1.9% of the total U.S. housing stock of 140 million units in that year. In line with our model, we categorize as national landlords those holding residential properties in at least two MSAs in any given year. In our regressions, "National–Landlord" is an indicator for MSAs with the presence of national landlords for at least 11 years, or more than half our sample period, 1996–2016. Table OA.1 in the Online Appendix reports the list of residential real estate companies used to build our national landlord indicator, together with the regions and states where they hold properties. As the table shows, the majority of national landlords own properties in at least two regions and several states.[20] In line with our model, we expect a higher propensity for national market housing risk to be priced in MSAs with significant presence of national landlords.

In our regressions, we also control for the potential diversification effect of second homes. For each MSA, we build the ratio of second homes to total housing units, which captures the percentage of housing units used as second homes by households from the same MSA or from a different MSA. During our sample period, this ratio is on average 4%. These households arguably hold more diversified housing portfolios than households owning just one home, which could potentially affect whether zip-code housing risk is priced. Using this ratio, we build "Second–Homes", an indicator for MSAs with a percentage of second homes above the sample top quartile.

The second home variable does not directly to control for the possibility that households of

[20] In a robustness test, we also define national landlords based on residential real estate companies that own residential properties in at least one east region and one west region, one east region and the Midwest, and one west region and the Midwest. As for the base measure, we focus on MSAs with national landlords for at least 11 years. This leads to three of the real estate firms in the main test not to be classified as national, but all three are operating in MSAs where there are other national residential real estate companies. Therefore, excluding these three firms from our analysis has no effect on our results.

any given MSA have a second home in a different MSA from the one where they have their main residence, which could also affect whether zip-code housing risk is priced. To mitigate the concern that this potential effect could bias our findings, our zip-code housing risk regressions control also for the natural logarithm of average MSA income. Arguably, wealthier households are more likely to own a second home. Finally, we control for the average MSA size with the natural logarithm of population. We use these basic MSA demographics (i.e., second-home indicator, log of income, and log of population) as controls also in our MSA and national housing risk pricing regressions. Arguably, whether households own a second home, how wealthy they are, and how big the MSA is could potentially affect the size of the rental market and the relative presence of local and national landlords. As we discuss, below, we find that our results are robust to controlling for additional MSA characteristics.

3.3 Descriptive Statistics

Table 1 Panel A presents summary statistics for yearly excess housing returns at the zip code, MSA, and national levels for the 9,093 zip codes across the 135 MSAs in our sample. In addition to cross–zip code and cross–MSA means and medians, we also compute the cross–sectional dispersion in these statistics to get a first look into heterogeneity in housing returns across local markets. The zip code–level statistics, except for cross–sectional dispersion, are computed in three steps. First, we obtain the time series statistic for each zip code. Second, within each MSA, we obtain the cross–sectional mean of these statistics. Finally, we obtain the average across MSAs. The cross–sectional dispersion measure is computed by first taking the time series mean for each zip code. Next, we compute the average within each respective MSA, and, finally, we obtain the cross–sectional standard deviation across MSAs. MSA–level statistics are computed by first taking the time–series average of the MSA–level variables, and then computing the statistics across all MSAs.

As expected, we find that zip code–level excess housing returns are more volatile than both MSA–level and national excess housing returns, with standard deviations of 6.44%, 5.18%, and 5.19% per year, respectively. The zip code–level average excess return (1.67%) is higher than both the average MSA (1.22%) and national housing market (1.15%) excess housing returns. This tentatively suggests that, in our sample, we observe a positive risk–return relation across aggregation levels.[21] The average yearly excess housing returns may, at first glance, seem low. This is because we only consider capital gains for housing and because the sample period includes the global financial crisis, when housing experienced extremely low returns (e.g., from 2007 to 2011, the Zillow U.S. housing market index had cumulative yearly excess returns of -27.4%). In Table OA.2 in the Online Appendix, we present the same statistics for the period from April 1996 to December 2007 and find that the average yearly excess housing returns across aggregation levels are almost twice as large compared to those for the full sample.[22]

[Table 1]

Table 1 Panel B reports summary statistics for homeownership, local and national landlord housing ownership, second home ownership, and control variables used in the study. Correlations are in Table OA.3. With the exception of the national landlord and second home variables, which are available for 134 of the 135 MSAs, most of the other variables are available for 125 MSAs. The number of zip codes variable is available for all 135 MSAs. The mean homeownership rate is 69%, ranging from 51% in Los Angeles, California, to 81% in Monroe, Michigan. The average (standard deviation) Local Landlord Housing Ownership and National Landlord Housing Ownership are 27% (7.1%) and 0.8% (0.7%), respectively. National landlord operate on average in XX% of the MSAs in our sample. Second homes constitutes 3.7% of all housing units during our sample period (Second Home Ownership), ranging from a minimum of 0.3% in ABC to a maximum of 37% in DEF. Average income and population are about $80,644 and 1.6 million, respectively. Overall, these descriptive statistics suggest that there is significant variation in homeownership rates, rental market structure (local and national landlords), and demographics across MSAs. In the next section, we first perform standard asset pricing tests to identify which sources of housing price risk – zip-code residual housing risk, MSA housing risk, and U.S. housing risk – are priced across MSAs. In line with our model, we next test whether the propensity for these risks to be priced depends on the relative presence of different types of housing investors, namely households, local landlords, and national landlords.

[21] Note that the averages are equally weighted across zip codes and across MSAs. Therefore, we cannot directly compare these averages based on disaggregate data to nationwide housing returns.

[22] Sagi (2021) includes capital expenditures when calculating commercial real estate returns. It would be desirable to include capital expenditures in the computation of housing returns. However, to the best of our knowledge, zip code–level data on capital expenditures at a monthly frequency is not available for residential real estate.

4 Main Results

The first step of our empirical analysis is to estimate the multifactor asset pricing model in Section 2.5 for each of the 135 MSAs in our sample. We therefore perform cross–sectional asset pricing tests for zip code housing returns within 135 different cross sections. As discussed in Section 2.5, our model includes three systematic risk factors: nation–wide housing market, stock market returns, and MSA–level housing returns. We also include zip code–specific residual volatility.

4.1 Factor Exposures and Zip Code Residual Housing Risk Estimates

We first examine the estimated zip code–level factor exposures and zip code–level residual volatilities, which are based on full–sample time series regressions based on Eq. (16) at the zip code level. Table 2 presents descriptive statistics of the risk exposures (betas) and the σ^{ZIP} estimates of the first–stage time series regressions. The table reports cross–sectional averages across all 9,093 zip codes.

[Table 2]

We find that the average U.S. stock market risk exposure β^{EQ} is essentially zero, with a mean (median) of -1.3e–3 (-8.9e–4). Even when considering lower–frequency (quarterly and annual) returns, we find that the vast majority of stock market betas are close to zero and statistically insignificant.[23] This finding suggests that home ownership at the zip–code level does not lead to exposure to U.S. stock market risk. Consequently, combining residential real estate with stocks can potentially provide diversification benefits. This finding is consistent with that of Case and Shiller (1989) and Jordà et al. (2019), who find a very low correlation between aggregate real estate returns and stock market returns. Notably, however, Cotter, Gabriel and Roll (2018) find declining diversification benefits in recent years across a range of assets, including equity, debt, and real estate, within and across countries.

As expected, the average national housing risk beta β^{US} is close to one, with a mean (median) of 1.05 (0.95). The beta does vary substantially across MSAs, as indicated by the cross–sectional standard deviation of 0.69. Further, we find that zip code–level housing returns carry exposure

[23]These results are discussed in Section OA.1 and reported in Figure OA.1 of the Online Appendix.

to local MSA–level housing risk. The mean and median β^{MSA} are 0.77 and 0.81, respectively. These beta estimates are less dispersed than the national housing risk betas, with a cross–sectional standard deviation of 0.40. These results suggest that it is important to allow for locally segmented housing markets, since local factors may be relevant. In the next section, we explicitly test whether local housing risk is priced.

Figure 1 plots a histogram of Newey and West (1994) adjusted t–statistics for the estimated risk exposures. We see that the U.S. stock market beta is generally insignificant across zip codes. Only 420 (220) of the 9,093 zip codes have positive and significant U.S. stock market betas at the 10% (5%) significance level. On the other hand, we find that the national and local housing risk betas are positive and significant for the vast majority of zip codes. Specifically, at the 10% (5%) level of significance, we find that 8,321 (8,106) and 7,877 (7,629) of the zip codes have positive and significant national and local housing risk betas, respectively. Further, only 20 (16) and 24 (18) of the national and local housing risk betas, respectively, are negative and significant at the 10% (5%) significance level, which, in line with asset pricing theory, suggests that beta estimates are almost always positive.

[Figure 1]

The monthly zip code-level residual housing risk (σ^{ZIP}) has a mean value of 0.66%. The cross–sectional standard deviation is 0.31%. To understand these values in relation to total risk, we also compute the full–sample standard deviation of zip code excess returns as a proxy for total risk (volatility). In comparison, the monthly total volatility has a mean of 0.93%. We compute the ratio of monthly zip code-level housing risk to total risk and find that, across zip codes, the average zip code-level residual housing risk represents 62.79% of total risk. This finding resonates partly with the stock market idiosyncratic volatility literature, which finds that stock volatility consists mostly of idiosyncratic volatility (e.g., Ang et al., 2009). Further, these results are consistent with Sagi (2021) and Giacoletti (2021), who find that idiosyncratic volatility constitutes the largest part of the return volatility of commercial and residential real estate returns, respectively. Systematic risk is also sizable representing roughly 37% of total risk.

4.2 Estimation of Prices of Risk

This variance decomposition discussed in the previous section does not necessarily imply that residual zip code–level housing risk is priced. Therefore, in the this section, we perform second–stage Fama–MacBeth cross–sectional regressions to examine which sources of risk drive expected housing returns at the zip code level. We estimate MSA–specific prices of risk for each risk factor by running the cross–sectional regression specified in Eq. (18) within each of the 135 MSAs in our estimation sample.

As shown in Appendix A.2, our model predicts that the prices of risk are all nonnegative. Accordingly, we choose to focus mainly on positive prices of risk for all our types of risk and perform single–sided tests.[24] Furthermore, in Section 4.3, we analyze the testable implications from our model that link cross–MSA variation in risk premiums to differences in ownership structures. Table 3 presents descriptive statistics for the estimated prices of risk. We report the number of MSAs that have significant positive prices of each risk at the 5% or 10% significance level, using Newey–West standard errors with four lags.[25] We also report the average price of risk and cross–sectional dispersion in the price of risk estimates among MSAs, where the respective price of risk is statistically significant.

Housing as an Investment Good Table 3, Panel A shows that, in 96 of our 135 MSAs—representing 76% of the total sample population and 76% of the total number of zip codes—at least one price of risk is positive and significant at the 10% significance level.[26] Thus, in the majority of MSAs, housing (at the zip code level) displays investment good properties, where homeowners receive compensation for bearing housing risk. Next, we consider which sources of risk are priced.

[Table 3]

[24]We note that the p–hacking effect discussed by Harvey et al. (2016) and Harvey (2017) for the case of stock returns is unlikely to be an issue in our setting, because the ZHVI database has hardly been used to study the cross section of housing returns and we use a parsimonious approach guided by economic theory to select our risk factors.

[25]The four lags are obtained following the formula of Newey and West (1994) for lag selection, $4(T/100)^{2/9}$, where T is the length of the time series (number of monthly observations). In our case, the average number of observations per zip code is 230, which yields four lags when rounding down. In untabulated results, we increase and decrease the number of lags, but this does not affect our results.

[26]At the 5% level, in 84 MSAs we find at least one significant and positive price of risk estimate.

National Housing Market Risk For the national housing market risk factor, we find that, of 135 estimated MSAs, 25 have positive and significant price of risk estimates at the 10% significance level, with an average price of risk of 0.15% per month and cross–MSA dispersion in the price of risk estimates of 0.13%.[27] At the 5% level, we find a significant price of risk in 20 MSAs. Thus, in about one out of five MSAs, the national housing market risk factor is positively priced. Interestingly, although aggregate house price indices are commonly used as indicators of the housing market for owner–occupiers, for the majority of them aggregate housing risk is not priced.

Local Housing Risk For the local MSA–level housing risk factor, we find that out of all 135 MSAs, 49 have positive and significant price of risk estimates at the 10% significance level, with an average price of risk of 0.13% per month and cross–sectional dispersion of 0.09%. Thus, at the 10% significance level, in more than a third of MSAs, local housing risk is significantly priced. At the 5% level the number decreases to 42 MSAs. Further, among these 49 (42) MSAs, we find that the average yearly risk premium (i.e., the average exposure multiplied by the price of risk estimate) is 1.27% (1.03%), which is significant compared to the average yearly excess return of 1.22% across all 135 MSAs, as reported in Table 1. Taken together, these results suggest that local MSA housing risk plays an important role in the pricing of the cross section of zip code–level housing returns. This finding is in line with earlier papers showing that housing markets cluster and that local aspects matter for housing returns (Goetzmann, Spiegel and Wachter, 1998; Tuzel and Zhang, 2017; Piazzesi, Schneider and Stroebel, 2020). Notably, there is also significant cross–MSA variation.

Zip Code–Level Housing Residual Risk Zip code–level housing residual risk carries a positive and significant risk premium in 48 of 135 MSAs (at the 10% significance level). The average price of risk is 0.10 per month and cross–MSA dispersion is 0.05. At the 5% level we find significant prices of risk in 32 MSAs. Further, among these 48 (32) MSAs, we find that the average yearly risk premium is 0.77% (0.58%). While this risk premium is smaller than the risk premium for the local housing risk, it is economically sizable. These results suggest that zip code specific housing risk plays an important role in the pricing of the cross section of zip code–level housing returns.

[27] Note that at the 10% significance level, we expect to find significant prices of risk in 14 out of 135 MSAs by chance alone. The observed 25 significant prices of risk exceeds this number.

This finding is in line with many homeowners being underdiversified and less able (or unable) to obtain a well–diversified housing portfolio.

U.S. Stock Market Risk In 30 MSAs, we find a significantly positive price of U.S. stock market risk at the 10% level. The average price of risk is 2.22% per month. While this is larger than the housing risk premiums, as reported in Table 2, the equity market betas are close to zero, indicating that the risk premium is also close to zero.[28] Taken together, these results suggest that U.S. stock market risk plays a non-significant role in the pricing of the cross section of zip code–level housing returns.[29]

Factor Pricing Patterns The results display important heterogeneity across MSAs in terms of which sources of risk are priced. To examine whether the significant prices of risk are clustered in a subset of MSAs or spread out, we next count how many MSAs have one or more sources of risk priced. Results are reported in Table 3, Panels B and C, respectively. Overall, we find that the MSAs where a given source of risk is priced do not necessarily overlap with the subset of MSAs where another source of risk is priced. 49 (31) MSAs have at least two significant prices of risk at the 10% (5%) level, which represents 41.0% (29.6%) of the total sample population and 42.7% (31.8%) of total number of zip codes. Only seven (five) MSAs have three or more significant prices of risk, and none of the MSAs have all four sources of risk significantly priced. In other words, we find substantial heterogeneity across local markets in terms of which sources of risk are priced.

To further investigate geographical patterns in the pricing of housing risk across the country, Figure 2 plots heatmaps of the U.S. MSAs where the price of risk estimates are positive and significant at the 10% significance level are marked in black and those where the price of risk estimate is not significant are marked in grey.[30] Figure 2a shows that national housing risk carries a significant positive price of risk mainly in a few MSAs on the East and West Coasts. In contrast, Figure 2b shows that local housing risk is priced in a broader set of MSAs across the U.S., including areas on the East Coast (e.g., Baltimore, Boston, and Washington), the West Coast (e.g., San Diego,

[28]The average risk premium among MSAs where U.S. stock market risk is significantly priced at the 10% (5%) significance level is just –0.06% (–0.03%) per year.

[29]In a robustness test in Table OA.4 of the Online Appendix, we exclude the U.S. stock market factor and find very similar results for the pricing of the different types of housing risk.

[30]Note that the plots do not mark all regions because our sample only covers 135 MSAs.

San Francisco, and Seattle), the Midwest (e.g., Chicago, Cleveland, and St. Louis), the Southwest (e.g., Albuquerque, Dallas, and San Antonio), and the Southeast (e.g., Charlotte, Memphis, and Nashville). We also see that local MSA housing risk is priced in smaller MSAs across the country, such as Cedar Rapids, Iowa; Kingston, New York; Lakeland, Florida; and Toledo, Ohio.

[Figure 2]

Figure 2c shows that zip code–specific housing residual risk is priced in a widespread range of MSAs across the country, including areas on the East Coast (e.g., Boston, New York, and Washington), the West Coast (e.g., San Diego, San Jose, and San Francisco), the Midwest (e.g., Cleveland, Columbus, and Minneapolis–St. Paul), the Southwest (e.g., Austin, Denver, and Houston), and the Southeast (e.g., Atlanta and Charlotte). Notably, zip code–specific housing risk is also priced in smaller MSAs, such as Akron, Ohio; Springfield, Illinois; and Utica, New York.

Finally, Figure 2d shows that U.S. stock market risk is significantly priced in many MSAs across the U.S. However, as shown in Table 2, the average beta is essentially zero, which means the risk premium is not economically sizable.

4.3 Risk Premia, Homeownership, and Rental Markets

Our analysis shows significant heterogeneity across MSAs in the pricing of the different sources of housing risk. Following our theoretical model, we now test whether this heterogeneity is related to the relative presence of different types of investors: (1) households, who have the least diversified housing portfolio because they typically own just one home in a single zip code, (2) local landlords, who own multiple housing units across zip codes within an MSA, and (3) national landlords, who own multiple housing units across MSAs. To perform our tests, we estimate three different probit models in which the dependent variables are indicators for whether zip code-level housing residual risk (λ^{ZIP}), MSA housing risk (λ^{MSA}), and national housing risk (λ^{US}) are positive and statistically significant at the 10% level or higher, respectively. The probit model is discussed in more detail in Section 2.5. Our main MSA-level explanatory variables include High–Homeownership, Local–Landlord, and National–Landlord. Our control variables include Second–Homes, and price to income ratio.

Table OA.5 in the Online Appendix reports the list of the 135 MSAs in our sample with their names and states. The table also reports homeownership rate, whether there is a significant presence of owner-occupiers, local and national landlord housing ownership, whether there is significant presence of local or national landlords, respectively, and indicators for whether zip-code level housing residual risk, local housing risk, and national housing risk are positive and statistically significant at the 10% level or higher, respectively. There are 32 MSAs with significant presence local landlords, and 28 MSAs with significant presence of national landlords.

Table 4 reports marginal effects from our probit model estimations.[31] In line with our model prediction, the evidence in column (1) suggests that zip code residual housing risk (λ^{ZIP}) is more likely to be priced in MSAs with high homeownership, our proxy for the presence of owner–occupants with underdiversified housing investments.[32] In column (2), specification with control variables, the coefficient on High-Homeownership (statistically significant at the 1% level) indicates that zip code residual housing risk is 29.4% more likely to be priced in MSAs with high homeownership (i.e., a significant presence of households with underdiversified housing portfolio). Although estimation of prices of risk depend on number of zip codes, this is not a concern in our case because we use an indicator for whether risk is priced in our main probit regressions. Nevertheless, in Table OA.6, we add number of zip codes as control and all our results hold.

[Table 4]

Column (2) also shows that the national landlord indicator enters the λ^{ZIP} regression with a statistically insignificant coefficient. In Table OA.7, we estimate the λ^{ZIP} regression in column (2) by adding the Local-Landlord variable, but without High-Homeownership, and we find that both the local landlord and national-landlord indicators are statistically insignificant. In the spirit of our theory, local and national landlords hold more diversified housing portfolios and, therefore, their presence should have little effect on whether zip code residual housing risk is priced. The coefficient of the second home indicator is negative and significant, suggesting that zip code residual risk is less likely to be priced if households hold more diversified housing portfolios.

[31] Results are similar if we focus on prices of risk that are positive and significant at the 5% level or higher. Further, results are qualitatively similar if we use continuous prices of risk (with insignificant prices of risk set equal to 0) and estimate tobit regressions. Refer to Table OA.8 in the Online Appendix. Further, as Table OA.9 in the Online Appendix shows, results are robust if we also control for unemployment and loan-to-value ratio.

[32] We find qualitatively very similar results if we use a continuous homeownership variable.

In columns (1) and (2), we measure the presence of underdiversified housing investors with owner occupancy. It is possible, however, that some small local landlords own housing units only within a zip code or a few neighboring zip codes. While this information is not readily available at the MSA level, we are able to proxy for the number housing units owned by small local landlords as follows. We obtain the number of small rental units (1-4 units) owned by individual investors from the Rental Housing Finance Survey and we assign these units to each MSA based on their population. We then add and subtract these units in our calculation of Homeownership and Local Landlord Housing Ownership, respectively, and build the corresponding adjusted High-Homeownership and Local-Landlord indicators. Refer to Table A.1 for more details on the construction of these variables. As Table OA.10 shows, our zip code residual housing risk, MSA housing risk, and national housing risk results hold using these adjusted measures.

Turning to MSA housing market risk (λ^{MSA}), Table 4, column (4), specification with control variables, shows that our local landlord indicator enters the estimations with a significantly positive coefficient. In line with the insights of our model, this finding suggests that local housing market risk is more likely to be priced if the MSA is characterized by a significant presence of landlords that operate only within that MSA. The coefficient estimate of 0.279, statistically signficant at the 5% level, for the Local–Landlord dummy suggests that local housing risk is 27.9% more likely to be priced in MSAs with significant presence of local landlords. Notably, the national landlord indicator and the other control variables are statistically insignificant in the MSA risk regression. In Table OA.7, we estimate the MSA housing risk regression with both the homeownership and the national landlord dummy (but without the local landlord indicators), and, in line with our theory, we find that these two variables are statistically insignificant.

Table 4, columns (5)-(7), report evidence from the national housing market risk (λ^{US}) estimations. Focusing on columns (6) and (7), we find that national housing market risk is more likely to be priced if there is a significant presence of national landlords. For example, the coefficient on National–Landlord in column (7) suggests that a significant presence on national landlords increases the propensity that national housing risk is priced by 18.1%, which is in line with our theoretical model. Further, as expected, High–Homeownership and Local–Landlord are statistically insignificant in columns (6) and (7), respectively. In Table OA.7, we estimate λ^{US} with both the

High–Homeownership and Local–Landlord variables, but without the national landlord indicator, and we find that both variables are statistically insignificant.

Overall, the evidence in Table 4 indicates that, in line with the implications of our theoretical model, the relative presence of different types of investors in the housing market is significantly related to the observed cross–MSA variation in the pricing of different types of housing risk. In areas with more locally constrained investors, local housing market risk is more likely priced.

5 Robustness Tests

5.1 Zillow Transaction and Assessment Data Price Indexes

In our first test, we check the robustness of our main results in Table 3, using our own ZTRAX–based zip code–level indexes, which covers ABC of the 135 MSAs in our main sample.[33] In this test, we still use the ZHVI MSA-level and national house indexes from our main specification as the local and national housing risk factors, respectively. The results are reported in Table 5.

[Table 5]

Panel A and B presents the descriptive statistics for the different prices of risk. Overall, results based on our own ZTRAX–based zip code–level indexes are similar to to our main findings in Table 3. That is, the majority of MSAs (80.1%) have at least one significant price of risk. Local housing market risk is the dominant risk factor, while zip code-level housing residual risk in this sample is still important yet less dominant than national housing market risk. Importantly, in line with the evidence in Table 4, in Panel C we find that the relative presence of different types of investors in the housing market is significantly related to observed the cross–MSA variation in the pricing of different types of housing risk. Namely, zip code–level housing residual risk, local housing market risk, and national housing market risk are more likely to be priced if homeownership is high, if there is significant presence of local landlords, and if there is significant presence of national landlords, respectively.

[33] We describe in detail the data and construction of these indexes in the Online Appendix.

5.2 Unsmoothing Housing Returns

As it is well documented, housing returns are persistent (see e.g. Fisher, Geltner and Webb, 1994; Geltner, 1991; Ross and Zisler, 1991; Geltner, 1993; Geltner, MacGregor and Schwann, 2003). In Table OA.11 of the Online Appendix, we report the median persistence of zip code–level, MSA–level, and U.S.–level housing returns. We can see strong persistence in housing returns at any aggregation level. Zip code (MSA) housing returns have a median persistence of 0.81 (0.79), while U.S. housing market returns have a persistence of 0.96. By comparison, monthly returns on the U.S. stock market have a persistence of 0.07 during our sample period.

To assess whether this persistence could bias out findings, in a robustness test, we use unsmoothed housing returns as a basis for our asset pricing tests. We find that an AR(3) specification that accounts for a monthly and a quarterly lag is enough to remove the persistence of housing returns at any aggregation level. Table 6, Panels A and B present descriptive statistics for the different prices of risk using unsmoothed housing returns. In line with out main results in Table 3, we find that the vast majority of MSAs (87.4%) have at least one significant price of risk estimate. Further, zip code-level housing residual risk and local housing market risk are the most important risks. Finally, in line with the evidence in Table 4, in Panel C we find again that the relative presence of different types of investors in the housing market is significantly related to the observed cross–MSA variation in the pricing of different types of housing risk.

[Table 6]

5.3 Subsample Analysis

To exclude the potential effect of the subprime crisis. in our next robustness test, we focus on period 1996–2007. As Table OA.12 in the Online Appendix shows, our main results and conclusions hold in this subsample. Notably, the number of national landlords increased from ABC during 1996-2010 to XYZ during 2011-2016. As expected, Table OA.13 of the Online Appendix shows that national housing risk is priced in a larger number of MSAs than in the full sample (38 versus 25) or the pre-Financial crisis subsample (38 versus 31).

5.4 Zip Code Requirement

In our main analysis, we require a minimum of 20 zip codes to estimate the price of risk within each MSA. As a robustness check, we estimate our model using a minimum of 15 zip codes per MSA, which leads to a larger estimation sample of 178 MSAs, versus 135 MSAs previously. The results of these estimations can be found in Table OA.14 of the Online Appendix. Our asset pricing and probit results are robust to this requirement.

5.5 Other Zip Code Characteristics

Our main model includes zip code–level housing volatility as the only zip code–level characteristic. To check the robustness of our main results, we include other zip code–level characteristics. These types of characteristics have been shown to matter for the cross–section of stock returns and may possibly matter for housing returns as well. In particular, we include a measure of value in the housing market and two different measures of size in our multi–factor asset pricing model. Following Asness, Moskowitz and Pedersen (2013), our proxy for value in the housing market is estimated as the negative value of the last 60 months' cumulative zip code returns. The first measure of size is the logarithm of the median zip code–level home's square footage. The second is the logarithm of the median zip code house price. These results can be found in Table OA.15 of the Online Appendix. Overall, we find that our main results and conclusions hold when these characteristics are included. Across all specifications, zip code-level housing residual risk and local housing are still the dominant risks.

6 Conclusion

This paper examines the effect of investor underdiversification on risk premia. We focus on housing, because constraints to diversification in the housing market are relatively explicit, yet they are also heterogenous across investors. The majority of U.S. households is homeowner[34] and most own only one property. At the same time, more diversified landlords and institutional investors are also present in housing markets. The heterogeneity across investors in their ability to hold a well–

[34]For instance, in the first quarter of 2022, 65.4% of U.S. households owned a house (U.S. Census Bureau, "Quarterly Residential Vacancies and Homeownership, First Quarter 2022").

diversified portfolio matters for asset pricing. We show, both theoretically and empirically, that the relative presence of diversified versus underdiversified investors is related to the types of risk that are priced in the cross–section of housing returns.

Using monthly housing returns of 9,093 zip codes across 135 different metropolitan areas (MSAs) in the U.S., we document three main results. First, we provide extensive evidence of a risk–return trade-off in housing returns at the zip code–level. In 71% of the 135 MSAs in our sample, at least one source of risk carries a significant positive risk premium. Hence, even at this highly disaggregated level, where many market participants are owner–occupiers, housing plays a dual role of a consumption and an investment good. Second, we find substantial heterogeneity across local markets in terms of which sources of risk are priced. Local MSA and zip code–level housing risk are most often priced; each in about half of the MSAs where housing has investment–good properties. In contrast, national housing market risk is priced in only 19% of the MSAs in our sample. Third, we show that the geographical variation in housing risk premia is related to the types of investors in the local housing markets. In line with our theoretical model, we show that in MSAs with high homeownership rates, zip code–level idiosyncratic housing risk is significantly more likely to be priced in terms of a higher risk premium, than in areas with low homeownership rates. Also, we find that in MSAs with a greater presence of local and national residential REITs, local MSA and national housing market risk are more likely priced.

Our findings have important implications for homeowners who are owner–occupants. While they face significant idiosyncratic housing risk (we find that 63% of total housing risk is zip code–specific idiosyncratic risk), they are not always compensated in terms of a higher risk premium. Unpriced risk can have welfare implications (e.g., Goetzmann and Kumar, 2003). Not only do owner–occupants have an undiversified housing portfolio, housing wealth typically is a major part of their overall wealth portfolio, further amplifying the concentration risk homeowners face. Furthermore, even though aggregate U.S. house price indexes are often used to assess housing market movements, for most owner–occupiers, national housing risk is not priced.

Our findings can also help inform the current policy debate on the effect of an increased presence of institutional and out–of–town investors on the housing market in the U.S. and worldwide. Their presence has been increasing in recent years, which could be related to technological advancements and a greater supply of houses that became available for institutional investors after the great financial crisis in 2007 (Mills et al., 2019). The presence of these investors has been shown to drive up house prices (e.g., Chinco and Mayer, 2016; Badarinza and Ramadorai, 2018; Mills et al., 2019; Favilukis and Van Nieuwerburgh, 2021). We show that systematic national and MSA–level housing market risks are more likely priced in areas with a greater presence of such investors. Our analysis is done from an unconditional perspective. An interesting question for future work is whether institutional investors will expand more to local areas where zip code–level housing risk is priced and if so, how this affects risk premia in a dynamic setting. At the same time, every investor in residential real estate arguably faces at least some degree of diversification constraints; substantially expanding a housing portfolio across areas requires substantial management efforts (e.g., maintaining the buildings) and not every area has as many suitable properties that are available for sale to institutional investors.

Appendix

A Model Derivations

A.1 Proofs of Propositions

First note that to derive the second line of the pricing equation in Expression (5), we use the Sherman-Morrison-Woodbury identity,

$$(A+B)^{-1} = A^{-1} + A^{-1}\left(-B^{-1} - A^{-1}\right)^{-1} A^{-1}, \tag{A.1}$$

to write

$$\begin{aligned}\left(\Sigma^{-1} - \frac{N_C}{N}\left(\Sigma^{-1} - \Sigma^{*-1}\right)\right)^{-1} &= \Sigma + \Sigma\left(\frac{N_C}{N}\left(\Sigma^{-1} - \Sigma^{*-1}\right) - \Sigma\right)^{-1}\Sigma \\ &= \Sigma + \Sigma\frac{N_C}{N}\left(\Sigma^{-1}\left(\Sigma^{-1} - \Sigma^{*-1}\right) - \frac{N_C}{N}I\right)^{-1} \\ &= \Sigma\left(I + \frac{N_C}{N}\left(\left(I - \Sigma^{*-1}\Sigma\right)^{-1} - \frac{N_C}{N}I\right)^{-1}\right).\end{aligned} \tag{A.2}$$

This allows us to rewrite the first line of Expression (5) to the second line.

Next, in Proposition 2.1 we express the pricing equation as a function of betas with respect to the national market returns. Below is the proof of this proposition.

Proof. **Proposition 2.1**

Pre-multiplying eq. (5) by α' leads to the following expression for the total market risk premium:

$$\mu_m - r = \frac{1}{\tau N}\left(M\sigma_m^2\left(1 + \frac{N_C}{N}\beta'\omega_*\right)\right), \tag{A.3}$$

where $\omega_* = \omega/M$. Substituting the expression for $1/(\tau N)$ from eq. (A.3) back into the pricing equation leads to

$$
\begin{aligned}
\mu - r\iota &= (\mu_m - r)\left(M\sigma_m^2\left(1+\frac{N_C}{N}\beta'\omega_*\right)\right)^{-1}\left(\Sigma S + \frac{N_C}{N}\Sigma\omega\right) \\
&= (\mu_m - r)\left(\sigma_m^2\left(1+\frac{N_C}{N}\beta'\omega_*\right)\right)^{-1}\left(\sigma_m^2\beta + \frac{N_C}{N}\Sigma\omega_*\right) \\
&= \frac{(\mu_m - r)/\sigma_m^2}{\left(1+\frac{N_C}{N}\beta'\omega_*\right)}\left(\sigma_m^2\beta + \frac{N_C}{N}\Sigma\omega_*\right).
\end{aligned}
\tag{A.4}
$$

It now follows that

$$
\begin{aligned}
\mu - r\iota &= \beta(\mu_m - r) + \frac{(\mu_m - r)/\sigma_m^2}{\left(1+\frac{N_C}{N}\beta'\omega_*\right)}\left(\sigma_m^2\beta - \sigma_m^2\beta\left(1+\frac{N_C}{N}\beta'\omega_*\right) + \frac{N_C}{N}\Sigma\omega_*\right) \\
&= \beta(\mu_m - r) + \frac{(\mu_m - r)/\sigma_m^2}{\left(1+\frac{N_C}{N}\beta'\omega_*\right)}\frac{N_C}{N}\left(\Sigma - \beta\sigma_m^2\beta'\right)\omega_* \\
&= \beta(\mu_m - r) + \chi\sigma_{SR}\left(\Sigma - \beta\sigma_m^2\beta'\right)\omega_*,
\end{aligned}
\tag{A.5}
$$

where we use the notation

$$
\begin{aligned}
\chi &= \frac{N_C}{N\left(1+\frac{N_C}{N}\beta'\omega_*\right)} \\
\sigma_{SR} &= \frac{\mu_m - r}{\sigma_m^2}.
\end{aligned}
\tag{A.6}
$$

We now have the required result. □

Proof. **Proposition 2.2**

Since we assume that Σ_ε^{nat} is block-diagonal, we may write

$$
\Sigma = \beta\sigma_m^2\beta' + G_A'\Sigma_{\varepsilon A}^{nat}G_A + G_B'\Sigma_{\varepsilon B}^{nat}G_B. \tag{A.7}
$$

This can then be written as

$$
\begin{aligned}
\Sigma = {} & \beta\sigma_m^2\beta' \\
& + G_A'\beta_{mA}^{orth}\sigma_{mA\_orth}^2\beta_{mA}^{orth\prime}G_A + G_A'\Sigma_{\varepsilon A}G_A \\
& + G_B'\beta_{mB}^{orth}\sigma_{mB\_orth}^2\beta_{mB}^{orth\prime}G_B + G_B'\Sigma_{\varepsilon B}G_B,
\end{aligned} \tag{A.8}
$$

where $\Sigma_{\varepsilon A}$ captures zip code–level residual housing risk in MSA A and a similar definition for $\Sigma_{\varepsilon B}$.

Inserting the expression for Σ into the national pricing equation (6) and selecting the rows corresponding to MSA A leads to

$$
\begin{aligned}
\mu_A - r\iota &= G_A(\mu - r\iota) \\
&= G_A\beta(\mu_m - r) + \chi\sigma_{SR}G_A \\
&\times \left(G_A'\beta_{mA}^{orth}\sigma_{mA\_orth}^2\beta_{mA}^{orth\prime}G_A + G_A'\Sigma_{\varepsilon A}G_A + G_B'\beta_{mB}^{orth}\sigma_{mB\_orth}^2\beta_{mB}^{orth\prime}G_B + G_B'\Sigma_{\varepsilon B}G_B\right)\omega_* \\
&= G_A\beta(\mu_m - r) + \chi\sigma_{SR}\beta_{mA}^{orth}\sigma_{mA\_orth}^2\beta_{mA}^{orth\prime}G_A\omega_* + \chi\sigma_{SR}\Sigma_{\varepsilon A}G_A\omega_*,
\end{aligned} \tag{A.9}
$$

since $G_AG_A' = I$. Note that here, ι is a 2×1 factor of ones.

We now define

$$
\Sigma_A^{*-1} = \Sigma_A^{-1} - \frac{N_{H_A}}{N_A}\left(\Sigma_A^{-1} - \Sigma_{H_A}^{-1}\right), \tag{A.10}
$$

which is a function of the inverse covariance matrices of landlords and households in A. Here, $N_A = N_{H_A} + N_{L_A}$, i.e., the total number of households and landlords in MSA A. Local market clearing for MSA A, i.e. $D_A = S_A$, implies

$$
\begin{aligned}
S_A &= \tau N_A\Sigma_A^{*-1}(\mu_A - r\iota) + \tau N_I G_A\Sigma^{-1}(\mu - r\iota) \\
\mu_A - r\iota &= \frac{1}{\tau N_A}\Sigma_A^* S_{adjA},
\end{aligned} \tag{A.11}
$$

where $S_A = G_AS$ is the total supply of assets in MSA A and

$$
S_{adjA} = S_A - \tau N_I G_A\Sigma^{-1}(\mu - r\iota) \tag{A.12}
$$

is the total supply adjusted for the demand from institutional investors for assets in MSA A.

Premultiplying with the local market portfolio weights of MSA A, α'_{mA}, leads to

$$\mu_{mA} - r = \frac{1}{\tau N_A} \alpha'_{mA} \Sigma^*_A S_{adjA}. \tag{A.13}$$

To explicitly include the MSA–level excess market returns in the pricing equation, note that we can combine the expressions for $1/\tau$ from eq. (A.3) and (A.13),

$$\begin{aligned} \frac{1}{\tau} &= N \left(M \sigma^2_m \left(1 + \frac{N_C}{N} \beta' \omega_* \right) \right)^{-1} (\mu_m - r) \\ \frac{1}{\tau} &= N_A \alpha'_{mA} \Sigma^*_A S_{adjA} (\mu_{mA} - r), \end{aligned} \tag{A.14}$$

to write σ_{SR} as a function of $\mu_{mA} - r$. We obtain

$$\sigma_{SR} N \left(M \left(1 + \frac{N_C}{N} \beta' \omega_* \right) \right)^{-1} = N_A \alpha'_{mA} \Sigma^*_A S_{adjA} (\mu_{mA} - r), \tag{A.15}$$

so that

$$\begin{aligned} \sigma_{SR} &= \frac{N_A}{N} \left(M \left(1 + \frac{N_C}{N} \beta' \omega_* \right) \right) \alpha'_{mA} \Sigma^*_A S_{adjA} (\mu_{mA} - r) \\ &= \frac{1}{\chi} \frac{N_C N_A}{N^2} M \alpha'_{mA} \Sigma^*_A S_{adjA} (\mu_{mA} - r) \\ &= \frac{1}{\chi} \frac{N_C N_A}{N^2} \frac{M}{M_A} \frac{(\mu_{mA} - r)}{\sigma^2_{mA,adj}} \\ &= \frac{1}{\chi} \frac{N_C N_A}{N^2} \frac{M}{M_A} \sigma_{SR,A\_adj}, \end{aligned} \tag{A.16}$$

where

$$\sigma^2_{mA,adj} = \alpha'_{mA} \Sigma^*_A S_{adjA} \frac{1}{M_A} \tag{A.17}$$

and

$$\sigma_{SR,A\_adj} = \frac{\mu_{mA} - r}{\sigma^2_{mA,adj}}. \tag{A.18}$$

This allows us to write the pricing equation (A.9) as

$$\begin{aligned}\mu_A - r\iota = {} & G_A\beta\left(\mu_m - r\right) \\ & + \frac{N_C}{N}\left(\beta_{mA}^{orth\prime}G_A\omega_*\right)\frac{\sigma^2_{mA\_orth}}{\sigma^2_{mA,adj}}\frac{M}{M_A}\frac{N_A}{N}\beta_{mA}^{orth}\left(\mu_{mA} - r\right) \\ & + \frac{N_C}{N}\sigma_{SR,A\_adj}\frac{M}{M_A}\frac{N_A}{N}\Sigma_{\varepsilon A}G_A\omega_*.\end{aligned} \tag{A.19}$$

This is the required result. □

A.2 Prices of Risk Are Nonnegative

This section of the appendix shows that the prices of risk in our model are all nonnegative. Recall that the prices of risk are given by

$$\begin{aligned}\lambda_{US} &= \left(\mu_m - r\right) \\ \lambda_{MSA\_A} &= \frac{N_C}{N}(\beta_{mA}^{orth\prime}G_A\omega_*)\frac{\sigma^2_{mA\_orth}}{\sigma^2_{mA,adj}}\frac{M}{M_A}\frac{N_A}{N}(\mu_{mA} - r) \\ \lambda_{ZIP\_A} &= \frac{N_C}{N}\sigma_{SR,A\_adj}\frac{M}{M_A}\frac{N_A}{N}\bar{\omega}_{A*}\end{aligned}$$

First, akin the CAPM, the price of U.S. housing market, risk is given by $(\mu_m - r)$ which is positive.

Next, we show that the local prices of risk are nonnegative as well. Note that we can combine the expressions for $\frac{1}{\tau}$ from global and local market clearing to write:

$$\chi\sigma_{SR} = \frac{N_C N_A}{N^2}\frac{M}{M_A}\sigma_{SR,A\_adj}$$

where

$$\begin{aligned}\chi &= \frac{N_C}{N\left(1 + \frac{N_C}{N}\beta'\omega_*\right)} \\ \sigma^2_{mA,adj} &= \alpha'_{mA}\Sigma^*_A S_{adjA}\frac{1}{M_A} \\ \sigma_{SR,A\_adj} &= \frac{(\mu_{mA} - r)}{\sigma^2_{mA,adj}}\end{aligned}$$

This allows us to rewrite the local prices of risk as:

$$
\begin{aligned}
\lambda_{MSA\_A} &= \frac{N_C}{N}(\beta_{mA}^{orth\prime} G_A \omega_*) \frac{\sigma^2_{mA\_orth}}{\sigma^2_{mA,adj}} \frac{M}{M_A} \frac{N_A}{N} (\mu_{mA} - r) \\
&= \frac{N_C}{N} \sigma_{SR,A\_adj} \frac{M}{M_A} \frac{N_A}{N} \left[\begin{array}{cc} cov(r_{A1}, r_{mA}^{orth}) & cov(r_{A2}, r_{mA}^{orth}) \end{array} \right] G_A \omega_* \\
&= \chi \sigma_{SR} \left[cov(r_{A1}, r_{mA}^{orth}) \omega_{*(1,1)} + cov(r_{A2}, r_{mA}^{orth}) \omega_{*(2,1)} \right] \\
\lambda_{ZIP\_A} &= \frac{N_C}{N} \sigma_{SR,A\_adj} \frac{M}{M_A} \frac{N_A}{N} \bar{\omega}_{A*} \\
&= \chi \sigma_{SR} \left[0.5 \omega_{*(1,1)} + 0.5 \omega_{*(2,1)} \right]
\end{aligned}
$$

This means that the prices of zip code and MSA risk differ only in terms of the weights given to the elements of ω_*. We assume that the zip code level returns positively covary with the orthogonalized local market return. Now, in order to show that the prices of local risk (λ_{MSA} and $_{ZIP}$) in the model are nonnegative it is sufficient to show that i) all elements of vector ω are nonnegative, and that ii) χ is nonnegative.

We start by showing i). First, recall that the pricing equation based on global market clearing can be expressed as

$$
\begin{aligned}
(\mu - r\iota) &= \frac{1}{\tau N} \Sigma S + \frac{1}{\tau N} \Sigma \frac{N_C}{N} \omega \\
&= \frac{1}{\tau N} \Sigma \left(D_C + D_I \right) + \frac{1}{\tau N} \Sigma \frac{N_C}{N} \omega
\end{aligned}
$$

where D_C and D_I are the demand from constrained and unconstrained (institutional) investors respectively. Pre-multiplying by $\tau N \Sigma^{-1}$ leads to

$$
\tau N \Sigma^{-1} (\mu - r\iota) = (D_C + D_I) + \frac{N_C}{N} \omega
$$

where the left hand side equals $\frac{N}{N_I} D_I$. This implies that we can now write ω as a function of the difference between the demand from unconstrained and constrained investors:

$$
\begin{aligned}
\frac{N}{N_I} D_I &= (D_C + D_I) + \frac{N_C}{N} \omega \\
\frac{N_C}{N_I} D_I - D_C &= \frac{N_C}{N} \omega
\end{aligned}
$$

So all elements of ω are positive if

$$\frac{N_C}{N_I}D_I - D_C \quad > \quad O$$

where O is a vector of zeros. In other words, to show that all elements of ω are positive we can also show that for each asset the scaled demand from unconstrained investors exceeds the demand from constrained investors:

$$\begin{aligned}
\frac{N_C}{N_I}D_I &> D_C \\
\frac{1}{N_I}D_I &> \frac{1}{N_C}D_C
\end{aligned}$$

In order to show this element by element, we take a simple two-asset case of one MSA A with two zip-codes. In this simple case, we only have constrained investors (i.e., households) and unconstrained investors (i.e., landlords and institutions who are now the same). The above condition can be written as

$$\begin{aligned}
\frac{1}{N_I}\tau N_I \Sigma_A^{-1}(\mu_A - r\iota) &> \frac{1}{N_C}\tau N_C \Sigma^{*-1}(\mu_A - r\iota) \\
\Sigma_A^{-1}(\mu_A - r\iota) &> \Sigma^{*-1}(\mu_A - r\iota) \\
(\mu_A - r\iota) &> \Sigma_A \Sigma^{*-1}(\mu_A - r\iota)
\end{aligned}$$

In a two-asset case we can easily write out the elements. Start by $\Sigma_A\Sigma^{*-1}$. We now have that

$$\Sigma^{*-1} = G_A' \Sigma_{H_A}^{-1} G_A$$

which allows us to write

$$
\begin{aligned}
\Sigma_A \Sigma^{*-1} &= \begin{bmatrix} \sigma^2_{A1} & \sigma_{A1A2} \\ \sigma_{A2A1} & \sigma^2_{A2} \end{bmatrix} \begin{bmatrix} \frac{N_{H_{A1}}}{N_C}(\sigma^2_{A1})^{-1} & 0 \\ 0 & \frac{N_{H_{A2}}}{N_C}(\sigma^2_{A2})^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \sigma^2_{A1}\frac{N_{H_{A1}}}{N_C}(\sigma^2_{A1})^{-1} & \sigma_{A1A2}\frac{N_{H_{A2}}}{N_C}(\sigma^2_{A2})^{-1} \\ \sigma_{A2A1}\frac{N_{H_{A1}}}{N_C}(\sigma^2_{A1})^{-1} & \sigma^2_{A2}\frac{N_{H_{A2}}}{N_C}(\sigma^2_{A2})^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \frac{N_{H_{A1}}}{N_C} & \sigma_{A1A2}\frac{N_{H_{A2}}}{N_C}(\sigma^2_{A2})^{-1} \\ \sigma_{A2A1}\frac{N_{H_{A1}}}{N_C}(\sigma^2_{A1})^{-1} & \frac{N_{H_{A2}}}{N_C} \end{bmatrix}
\end{aligned}
$$

We can now write out both elements of

$$
\begin{aligned}
(\mu_A - r\iota) &> \Sigma_A \Sigma^{*-1}(\mu_A - r\iota) \\
(\mu_A - r\iota) &> \begin{bmatrix} \frac{N_{H_{A1}}}{N_C} & \sigma_{A1A2}\frac{N_{H_{A2}}}{N_C}(\sigma^2_{A2})^{-1} \\ \sigma_{A2A1}\frac{N_{H_{A1}}}{N_C}(\sigma^2_{A1})^{-1} & \frac{N_{H_{A2}}}{N_C} \end{bmatrix} (\mu_A - r\iota) \\
\begin{bmatrix} (\mu_{A1} - r) \\ (\mu_{A2} - r) \end{bmatrix} &> \begin{bmatrix} \frac{N_{H_{A1}}}{N_C}(\mu_{A1} - r) + \sigma_{A1A2}\frac{N_{H_{A2}}}{N_C}(\sigma^2_{A2})^{-1}(\mu_{A2} - r) \\ \sigma_{A2A1}\frac{N_{H_{A1}}}{N_C}(\sigma^2_{A1})^{-1}(\mu_{A1} - r) + \frac{N_{H_{A2}}}{N_C}(\mu_{A2} - r) \end{bmatrix}
\end{aligned}
$$

To show that both elements of ω are positive in this two-asset case, we need to show that both inequalities given above hold. To this end, we make one assumption that $\sigma_{A2A1} > 0$, which implies positively correlated local housing markets.

So we have for element (1,1):

$$
\begin{aligned}
(\mu_{A1} - r) &> \frac{N_{H_{A1}}}{N_C}(\mu_{A1} - r) + \sigma_{A1A2}\frac{N_{H_{A2}}}{N_C}(\sigma^2_{A2})^{-1}(\mu_{A2} - r) \\
\frac{N_{H_{A2}}}{N_C}(\mu_{A1} - r) &> \sigma_{A1A2}\frac{N_{H_{A2}}}{N_C}(\sigma^2_{A2})^{-1}(\mu_{A2} - r) \\
\frac{(\mu_{A1} - r)}{(\mu_{A2} - r)} &> \sigma_{A1A2}(\sigma^2_{A2})^{-1}
\end{aligned}
$$

And for element (2,1):

$$
\begin{aligned}
(\mu_{A2} - r) &> \sigma_{A2A1}\frac{N_{H_{A1}}}{N_C}(\sigma^2_{A1})^{-1}(\mu_{A1} - r) + \frac{N_{H_{A2}}}{N_C}(\mu_{A2} - r) \\
\frac{N_{H_{A1}}}{N_C}(\mu_{A2} - r) &> \sigma_{A2A1}\frac{N_{H_{A1}}}{N_C}(\sigma^2_{A1})^{-1}(\mu_{A1} - r) \\
\frac{(\mu_{A2} - r)}{(\mu_{A1} - r)} &> \sigma_{A2A1}(\sigma^2_{A1})^{-1} \\
\text{assume } \sigma_{A2A1} &> 0: \\
\frac{(\mu_{A1} - r)}{(\mu_{A2} - r)} &< \sigma^{-1}_{A2A1}\sigma^2_{A1}
\end{aligned}
$$

For these two elements to hold, we must have at least consistency, so we must have that

$$
\sigma_{A1A2}(\sigma^2_{A2})^{-1} < \frac{(\mu_{A1} - r)}{(\mu_{A2} - r)} < \sigma^{-1}_{A2A1}\sigma^2_{A1}
$$

For this to hold, we must at least have that

$$
\begin{aligned}
\sigma_{A1A2}(\sigma^2_{A2})^{-1} &< \sigma^{-1}_{A2A1}\sigma^2_{A1} \\
\frac{\sigma^2_{A1A2}}{\sigma^2_{A2}\sigma^2_{A1}} &< 1
\end{aligned}
$$

which is the case given the left hand side equals the squared correlation between the two zip code returns. This implies that at least one of the two inqualities must hold, since if they both don't hold we would have that $\frac{\sigma^2_{A1A2}}{\sigma^2_{A2}\sigma^2_{A1}} > 1$, which cannot be true.

The conditions for both elements are given by:

$$
\begin{aligned}
\frac{(\mu_{A1} - r)}{(\mu_{A2} - r)} &> \sigma_{A1A2}(\sigma^2_{A2})^{-1} \\
\frac{(\mu_{A1} - r)}{(\mu_{A2} - r)} &< \sigma^{-1}_{A2A1}\sigma^2_{A1}
\end{aligned}
$$

We already know at least one should hold. Suppose the top one holds and given the assumption that $\sigma_{A1A2} > 0$, this implies for the bottom expression that

$$\begin{aligned}
\frac{(\mu_{A1} - r)}{(\mu_{A2} - r)} &< \sigma_{A2A1}^{-1}\sigma_{A1}^2 \\
\frac{(\mu_{A1} - r)}{(\mu_{A2} - r)}\sigma_{A2}^2 &< \sigma_{A2A1}^{-1}\sigma_{A2}^2\sigma_{A1}^2? \\
\frac{(\mu_{A1} - r)}{(\mu_{A2} - r)}\frac{\sigma_{A2}^2}{\sigma_{A2A1}} &< \frac{\sigma_{A2}^2\sigma_{A1}^2}{\sigma_{A2A1}^2} \\
\frac{(\mu_{A2} - r)}{(\mu_{A1} - r)}\frac{\sigma_{A2A1}}{\sigma_{A2}^2} &> \frac{\sigma_{A2A1}^2}{\sigma_{A2}^2\sigma_{A1}^2} \\
\beta_{A1,A2}(\mu_{A2} - r) &> \rho_{A1,A2}^2(\mu_{A1} - r)
\end{aligned}$$

where we can write

$$(\mu_{A1} - r) = c_{A1,A2} + \beta_{A1,A2}(\mu_{A2} - r)$$

which is based on an OLS regression:

$$(R_{A1,t} - R_{f,t}) = c_{A1,A2} + \beta_{A1,A2}(R_{A2,t} - R_{f,t}) + \varepsilon_{A1,t}.$$

We know from the top condition (element 1,1) that

$$\begin{aligned}
\frac{(\mu_{A1} - r)}{(\mu_{A2} - r)} &> \sigma_{A1A2}(\sigma_{A2}^2)^{-1} \\
(\mu_{A1} - r) &> \beta_{A1,A2}(\mu_{A2} - r) \\
c_{A1,A2} &> 0
\end{aligned}$$

But this does not tell us whether

$$\beta_{A1,A2}(\mu_{A2} - r) > \rho_{A1,A2}^2(\mu_{A1} - r)?$$

From element (1,1) we have

$$\begin{aligned}
(\mu_{A1} - r) &> \beta_{A1,A2}(\mu_{A2} - r) \\
(\mu_{A1} - r) &> \beta_{A1,A2}(\mu_{A2} - r) > \rho_{A1,A2}^2(\mu_{A1} - r)
\end{aligned}$$

So do we have

$$
\begin{aligned}
(\mu_{A1} - r) &> \rho^2_{A1,A2}(\mu_{A1} - r)? \\
1 &> \rho^2_{A1,A2} \\
&\text{so} \\
\beta_{A1,A2}(\mu_{A2} - r) &> \rho^2_{A1,A2}(\mu_{A1} - r)
\end{aligned}
$$

Indeed, which completes the derivation that both elements of ω are nonnegative.

To show that the prices of local risk are nonnegative, what remains to show is that $\chi > 0$. Recall that

$$
\chi = \frac{N_C}{N\left(1 + \frac{N_C}{N}\beta'\omega_*\right)}
$$

where

$$
\begin{aligned}
\omega_* &= K\alpha \\
\omega &= KS \\
K &= \left[(I - \Sigma^{*-1}\Sigma)^{-1} - \frac{N_C}{N} I\right]^{-1}
\end{aligned}
$$

We can write

$$
\begin{aligned}
\beta'\omega_* = \frac{1}{\sigma_m^2}\Sigma\alpha'\omega_* \\
= \frac{1}{\sigma_m^2}\Sigma\alpha' K\alpha > 0
\end{aligned}
$$

since K is an invertible matrix. This completes the proof that the elements in ω and χ are nonnegative, implying that the prices of local housing market risk in our model are nonnegative as well.

A.3 How Local Risk Premia Depend on the Presence of Underdiversified Investors

Recall from the previous section in the Appendix that the prices of local risk are similar, except for the weighting of the elements of ω_*:

$$\lambda_{MSA\_A} = \chi\sigma_{SR}\left[cov(r_{A1}, r_{mA}^{orth})\omega_{*(1,1)} + cov(r_{A2}, r_{mA}^{orth})\omega_{*(2,1)}\right]$$
$$\lambda_{ZIP\_A} = \chi\sigma_{SR}\left[0.5\omega_{*(1,1)} + 0.5\omega_{*(2,1)}\right]$$

To show the effect of having more households relative to landlords on the prices of MSA and zip code–level risk, we show the effect on the elements of ω_*. When these elements increase in magnitude, the effect will be stronger for λ_{ZIP_A} than for λ_{MSA_A} as the weights have larger magnitudes (i.e, equal weights of 0.5 exceed the covariance-based weights, which are of magnitude returns–squared). Let's compare ω with and without households, keeping the total number of constrained investors N_C and the number of unconstrained institutional investors N_I constant. Recall from the previous subsection A.2 that

$$\omega \quad = \quad \frac{N}{N_I}D_I - \frac{N}{N_C}D_C$$

With households (and landlords) as constrained investors, this can be written as

$$\omega_{withHH} \quad = \quad N\left(X_I - \frac{N_L}{N_C}X_{LL} - \frac{N_H}{N_C}X_{HH}\right)$$

Without households (having only landlords as constrained investors), this expression becomes

$$\omega_{noHH} \quad = \quad N\left(X_I - X_{LL}\right)$$

where X_i denotes the optimal portfolio weights of investor group i.

Next, we compare ω with and without households, keeping N_C constant:

$$\begin{aligned}\omega_{withHH} - \omega_{noHH} &= N\left(X_I - \tfrac{N_L}{N_C}X_{LL} - \tfrac{N_H}{N_C}X_{HH}\right) - N\left(X_I - X_{LL}\right) \\ &= N\tfrac{N_H}{N_C}\left(X_{LL} - X_{HH}\right)\end{aligned}$$

The above expression shows that the difference in ω with and without households is increasing in the number of households relative to the number of constrained investors. Hence, if there are more households, ω increases and the effect will be stronger for the price of zip code–level risk λ_{ZIP} as there the elements in ω are multiplied by equal weights of 0.5 instead of covariance-based weights.

In sum, we have shown that the prices of local MSA and zip code level risk are increasing in the total number of constrained investors. In additionl when the number of households relative to all constrained investors is higher, the effect on the price of zip code–level risk exceeds that on the price of MSA level risk. We test these implications in our empirical analysis, where we link the price of risk estimates to the relative presence of the different types of (un)constrained investors.

B Further Details about Zillow Data

In Table A.1, we report the list of all variables used in the paper including their data source, frequency, sample period and definition.

[Table A.1]

References

Adelino, M., Schoar, A., and Severino, F. (2016), "Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class," *Review of Financial Studies,* 29(7), 1635–1670.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006), "The Cross Section of Volatility and Expected Returns," *Journal of Finance,* 11(1), 259–299.

— (2009), "High Idiosyncratic Volatility and Low Returns: International and Further U.S. Evidence," *Journal of Financial Economics,* 91(1), 1–23.

Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013), "Value and Momentum Everywhere," *Journal of Finance,* 68(3), 929 – 985.

Badarinza, C., and Ramadorai, T. (2018), "Home away from home? Foreign demand and London house prices," *Journal of Financial Economics,* 130(3), 532–555.

Baldauf, M., Garlappi, L., and Yannelis, C. (2020), "Does Climate Change Affect Real Estate Prices? Only If You Believe In It," *Review of Financial Studies,* 33(3), 1256–1295.

Bekaert, G., Hodrick, Robert J., and Zhang, Xiaoyan (2009), "International stock return comovements," *The Journal of Finance,* 64(6), 2591–2626.

Bernstein, A., Gustafson, M. T., and Lewis, R. (2019), "Disaster on the Horizon: The Price Effect of Sea Level Rise," *Journal of Financial Economics,* 134(2), 253–272.

Blume, M.E., and Friend, I. (1975), "The Asset Structure of Individual Portfolios and some Implications for Utility Functions," *Journal of Finance,* 30(2), 585–603.

Cannon, S., Miller, N. G., and Pandher, G. S. (2006), "Risk and Return in the U.S. Housing Market: A Cross-Sectional Asset-Pricing Approach," *Real Estate Economics,* 34(4), 519–552.

Case, K. E., and Shiller, R. J. (1989), "The Efficiency of the Market of Single-Family Homes," *American Economic Review,* 79(1), 125–137.

Case, K. E., Cotter, J., and Gabriel, S. (2011), "Housing Risk and Return: Evidence from a Housing Asset-Pricing Model," *Journal of Portfolio Management,* 37(5), 89–109.

Chinco, A., and Mayer, C. (2016), "Misinformed Speculators and Mispricing in the Housing Market," *Review of Financial Studies,* 29(2), 486–522.

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

Conrad, J., Dittmar, R. F., and Ghysels, E. (2013), "Ex Ante Skewness and Expected Stock Returns," *Journal of Finance,* 68 (1), 85–124.

Cotter, J., Gabriel, S., and Roll, R. (2015), "Can Housing Risk be Diversified? A Cautionary Tale from the Housing Boom and Bust," *Review of Financial Studies,* 28(3), 913–936.

— (2018), "Nowhere to Run, Nowhere to Hide: Asset Diversification in a Flat World," *Working Paper.*

De Jong, F., and De Roon, F.A. (2005), "Time-varying market integration and expected returns in emerging markets," *Journal of Financial Economics,* 78, 583–613.

Della Preite, M., Uppal, R., Zaffaroni, P., and Zviadadze, I. (2022), "What is Missing in Asset-Pricing Factor Models?," *Working paper, Imperial College, EDHEC and HEC Paris.*

Dimmock, S., Kouwenberg, R., Mitchell, O., and Peijnenburg, K. (2020), "Household Portfolio Underdiversification and Probability Weighting: Evidence from the Field," *The Review of Financial Studies,* 34(9), 4524–4563.

Eichholtz, P., Korevaar, M., Lindenthal, T., and Tallec, R. (2021), "The Total Return and Risk to Residential Real Estate ," *Review of Financial Studies,* 34(8), 3608–3646.

Errunza, V., and Losq, E. (1985), "International Asset Pricing under Mild Segmentation: Theory and Test," *Journal of Finance,* 40(1), 105–124.

Fama, E. F., and MacBeth, J. D. (1973), "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy,* 81(3), 607–636.

Favilukis, J., and Van Nieuwerburgh, S. (2021), "Out-of-Town Home Buyers and City Welfare," *Journal of Finance,* 76(5), 2577–2638.

Fisher, J. D., Geltner, D. M, and Webb, B. R. (1994), "Value Indices of Commercial Real Estate: A Comparison of Index Construction Methods," *Journal of Real Estate Finance and Economics,* 9(2), 137–164.

Flavin, M., and Yamashita, T. (2002), "Owner-occupied housing and the composition of the household portfolio," *The American Economic Review,* 92(1), 345–362.

French, K., and Poterba, J. (1991), "International Diversification and International Equity Markets," *American Economic Review,* 81, 222–226.

Geltner, D. M. (1991), "Smoothing in Appraisal-based Returns," *Journal of Real Estate Finance and Economics,* 4, 327–345.

— (1993), "Estimating Market Values from Appraisal Values without Assuming an Efficient Market," *Journal of Real Estate Research,* 8, 352–345.

Geltner, D. M., MacGregor, B. D., and Schwann, G. M. (2003), "Appraisal Smoothing and Price Discovery in Real Estate Markets," *Journal of Real Estate Finance and Economics,* 9(5-6), 1047–1064.

Genesove, D., and Han, L. (2007), "Search and Matching in the Housing Market," *Journal of Urban Economics,* 72(1), 31–45.

Ghent, A. (2021), "What's Wrong with Pittsburgh? Delegated Investors and Liquidity Concentration," *Journal of Financial Economics,* 139(2), 337–358.

Giacoletti, M. (2021), "Idiosyncratic Risk in Housing Markets," *Review of Financial Studies,* 34(8), 3695–3741.

Giglio, S., Maggiori, M., Rao, K., Stroebel, J., and Weber, A. (2021), "Climate Change and Long-Run Discount Rates: Evidence from Real Estate," *Review of Financial Studies,* 34(8), 3527–3571.

Goetzmann, W., and Kumar, A. (2003), "Diversification Decisions of Individual Investors and Asset Prices," *Working paper, Yale University and University of Notre Dame.*

— (2008), "Equity Portfolio Diversification," *Review of Finance,* 12, 433–563.

Goetzmann, W. N., Spiegel, M., and Wachter, S. M. (1998), "Do cities and suburbs cluster?" *Cityscape,* 3(3), 193–203.

Han, L. (2013), "Understanding the Puzzling Risk-Return Relationship for Housing," *Review of Financial Studies,* 26(4), 877–928.

Han, L., and Strange, W. C. (2014), "Bidding Wars for Houses," *Real Estate Economics,* 42(1), 1–32.

Harvey, C. R. (2017), "Presidential Adress: The Scientific Outlook in Financial Economics," *Journal of Finance,* 72(4), 1399–1440.

Harvey, C. R., Liu, Y., and Zhu, H. (2016), "... and the Cross-Section of Expected Returns," *Review of Financial Studies,* 29(1), 5–68.

Herskovic, B., Kelly, B., Lustig, H., and Van Nieuwerburgh, S. (2016), "The Common Factor in Idiosyncratic Volatility: Quantitative Asset Pricing Implications," *Journal of Financial Economics,* 119 (2), 249–283.

Huang, W., Liu, Q., Rhee, S. G., and Zhang, L. (2010), "Return Reversals, Idiosyncratic Risk, and Expected Returns," *Review of Financial Studies,* 23(1), 147–168.

Iachan, F., D., Silva., and Zi, C. (2022), "Under-Diversification and Idiosyncratic Risk Externalities," *Journal of Financial Economics,* 143(3), 1227–1250.

Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M., and Taylor, A.M. (2019), "The Rate of Return on Everything, 1870–2015," *The Quarterly Journal of Economics,* 134(3), 1225–1298.

Kelly, M. (1995), "All Their Eggs in One Basket: Portfolio Diversification of US Households," *Journal of Economic Behavior and Organization,* 27(1), 87–96.

Koijen, Ralph S. J., and Yogo, Motohiro (2019), "A Demand System Approach to Asset Pricing," *Journal of Political Economy,* 127(4), 1475–1515.

Landvoigt, T., Piazzesi, M., and Schneider, M. (2015), "The Housing Market(s) of San Diego," *American Economic Review,* 105(4), 1371–1407.

Levy, H. (1978), "Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio," *American Economic Review,* 68 (4), 643–658.

Malkiel, B. G., and Xu, Y. (2004), "Idiosyncratic Risk and Security Returns," *Working Paper.*

Manson, S., Schroeder, J., Van Riper, D., Kugler, T., and Ruggles, S. (2022), "IPUMS National Historical Geographic Information System: Version 17.0 [dataset]," *Minneapolis: University of Minnesota.*

Massa, M., and Simonov, A. (2006), "Hedging, Familiarity and Portfolio Choice," *The Review of Financial Studies,* 19(2), 633–685.

Merton, R. C. (1987), "A Simple Model of Capital Market Equilibrium with Incomplete Information," *Journal of Finance,* 42 (3), 483–510.

Mian, A., and Sufi, A. (2009), "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *Quarterly Journal of Economics,* 124(4), 1449–1496.

Mian, A., Sufi, A., and Trebbi, F. (2015), "Foreclosures, House Prices, and the Real Economy," *Journal of Finance,* 70(6), 2587–2634.

Mills, J., Molloy, R., and Zarutskie, R. (2019), "Large-Scale Buy-to-Rent Investors in the Single-Family Housing Market: The Emergence of a New Asset Class," *Review of Financial Studies,* 47(2), 399–430.

Mitton, T., and Vorkink, K. (2007), "Equilibrium Underdiversification and the Preference for Skewness," *The Review of Financial Studies,* 20(4), 1255–1288.

Murfin, J., and Spiegel, M. (2020), "Is the Risk of Sea Level Rise Capitalized in Residential Real Estate?," *Review of Financial Studies,* 33(3), 1217–1255.

Newey, W.K., and West, K.D. (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies,* 61, 631–653.

Peng, L., and Thibodeau, T. G. (2017), "Idiosyncratic Risk of House Prices: Evidence from 26 Million Home Sales," *Real Estate Economics,* 45(2), 340–375.

Peng, L., and Zhang, L. (2021), "House Prices and Systematic Risk: Evidence from Microdata," *Real Estate Economics,* 49(4), 1069–1092.

Piazzesi, M., Schneider, M., and Stroebel, J. (2020), "Segmented Housing Search," *American Economic Review,* 110(3), 720–759.

Plazzi, A., Torous, W., and Valkanov, R. (2008), "The Cross-Sectional Dispersion of Commercial Real Estate Returns and Rent Growth: Time Variation and Economic Fluctuations," *Real Estate Economics,* 36(3), 403–439.

Ross, S. A., and Zisler, R. C. (1991), "Risk and Return in Real Estate," *Journal of Real Estate Finance and Economics,* 4, 175–190.

Ruggles, S., Genadek, K., Goeken, R., Grover, J., and Sobek, M. (2017), "IPUMS USA: Version 7.0 [dataset]," *Minneapolis: University of Minnesota.*

Sagi, S. J. (2021), "Asset-Level Risk and Return in Real Estate Investments," *Review of Financial Studies,* 34(8), 3647–3694.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2015), "Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle," *Journal of Finance,* 70(5), 1903–1948.

Tesar, L., and Werner, I. (1998), "The Internationalization of Securities Markets since the 1987 Crash," *Brookings-Wharton Papers on Financial Services, The Brookings Institution.*

Tuzel, S., and Zhang, M. B. (2017), "Local Risk, Local Factors, and Asset Prices," *Journal of Finance,* 72(1), 325–370.

Van Nieuwerburgh, S., and Veldkamp, L. (2010), "Information Acquisition and Under-Diversification," *The Review of Economic Studies,* 77(2), 779–805.

Van Nieuwerburgh, S., and Weill, P-O. (2010), "Why Has House Price Dispersion Gone Up?," *Review of Economic Studies,* 77(4), 1567–1606.

Warnock, F.E. (2002), "Home Bias and High Turnover Reconsidered," *Journal of International Money and Finance,* 21, 798–805.

Table 1: Descriptive Statistics Excess Returns

In Panel A, this table reports descriptive statistics including mean, median, minimum, maximum, and standard deviation (SD), for zip code, metropolitan statistical area (MSA), U.S. housing market and U.S. stock market yearly excess returns. We also report the cross–sectional dispersion (Dispersion) for zip code and MSA yearly excess returns. In Panel B, the table reports these statistics for MSA–level homeownership, local and national landlord housing ownership, the High–Homeownership, Local–Landlord, and National–Landlord indicators, respectively, second home ownership, the Second–Homes indicator, price to income ratio and number of zip codes in hundreds. All zip code statistics, except the cross–sectional dispersion, are computed in three steps. First, we take the time–series statistic for each zip code. Second, within each MSA we take the cross–sectional mean of these statistics. Finally, we take the average across MSAs. The cross–sectional dispersion is computed by first taking the time–series mean for each zip code, then taking the average within each respective MSA, and finally the cross–sectional standard deviation across MSAs. The MSA excess returns dispersion is computed by first taking the time–series mean of each MSA, and then taking the standard deviation of these means. The MSA–level statistics in Panel B are computed in two steps. First, we take the time–series average for each quantity. Second, we calculate each statistic across MSAs. The mean, median, minimum, maximum, and SD of all excess returns are annualized and are presented in percentage units. Refer to Table A.1 of the main text Appendix for detailed variable definitions. The sample period is April 1996 – December 2016.

	Panel A: Excess Returns (% pa)						
	N	Mean	Median	Minimum	Maximum	SD	Dispersion
Zip Code	9093	1.67	1.21	-30.17	39.71	6.44	1.85
MSA	135	1.22	0.61	-12.43	18.76	5.18	1.39
U.S. Housing Market		1.15	1.62	-8.66	9.35	5.19	
U.S. Stock Market		7.95	11.70	-38.34	35.20	18.75	

	Panel B: Ownership and MSA Fundamentals Data					
	N	Mean	Median	Minimum	Maximum	SD
Homeownership	125	0.684	0.692	0.508	0.812	0.052
High-Homeownership	125	0.248	0.000	0.000	1.000	0.434
Local Landlord Housing Own.	125	0.270	0.273	0.022	0.465	0.071
Local-Landlord	125	0.256	0.000	0.000	1.000	0.438
National Landlord Housing Own.	134	0.008	0.005	0.000	0.034	0.007
National-Landlord	134	0.254	0.000	0.000	1.000	0.437
Second Home Ownership	134	0.037	0.017	0.003	0.370	0.057
Second-Homes	134	0.254	0.000	0.000	1.000	0.437
Price to Income Ratio	125	2.583	2.369	1.200	6.269	0.877
Number of Zip Codes	135	67.356	37.000	20.000	736.000	84.987

Table 2: Factor Exposures and Zip Code-Level Housing Residual Risk

This table reports descriptive statistics of risk exposures and idiosyncratic housing risk for 9,093 zip codes across 135 MSAs. We report the mean, median, minimum, maximum, and standard deviation (SD) for the exposures to the U.S. stock market risk (β^{EQ}), national housing risk (β^{US}), local housing risk (β^{MSA}). We also report these statistics for zip code-level housing residual risk (ZIP). These are all estimated for each zip code using the three–factor model specified in Eq. (16). A detailed list of all variables can be found in Table A.1 of the Appendix. We also report these statistics for the total volatility (Total Vol.), which is measured by the full sample standard deviation, and for the ratio of ZIP to the standard deviation ($\overline{ZIP}$/Total Vol.). The mean, median, minimum, maximum, and SD of ZIP, total volatility, and $\overline{ZIP}$/Total Vol. are in monthly percentage units. The sample period is April 1996 – December 2016.

	Mean	Median	Minimum	Maximum	SD
$\beta^{\overline{EQ}}$	-1.3e-3	-8.9e-4	-0.15	0.12	0.01
$\beta^{\overline{US}}$	1.05	0.95	-7.46	5.43	0.69
$\beta^{\overline{MSA}}$	0.77	0.81	-2.87	5.70	0.40
$\overline{ZIP}$ (% per month)	0.66	0.58	0.17	2.41	0.31
Total Vol. (% per month)	0.93	0.86	0.26	2.53	0.33
$\overline{ZIP}$/Total Vol. (% per month)	62.79	62.18	10.67	99.99	19.70

Table 3: Prices of Risk

This table reports descriptive statistics of estimated prices of risk that are positive and statistically significant, based on single–sided tests. We estimate prices of risk for the national housing risk (λ^{US}), local housing risk (λ^{MSA}) and zip code-level housing residual risk (λ^{ZIP}) for each MSA with at least 20 zip codes. In Panel A, we report the number of estimated MSAs (N) and the number of MSAs where each price of risk is positive and significant ($\lambda > 0$). We also report average λ and the cross–sectional dispersion in λ (Dispersion) within the MSAs where each respective price of risk is positive and significant. In Panel B and C, we report the number of MSAs where different sets of prices of risk are significant: e.g., we report in which MSAs λ^{MSA} and λ^{ZIP} are positive and significant. We also report the share of the full sample population (Share of Population) and number of zip codes (Share of Zip Codes) these MSAs represent. Finally, we look at how many MSAs have one or more significant prices of risk. Significance is at the 5% or 10% level single–sided using Newey and West (1994) standard errors with 4 lags. The sample period is April 1996 – December 2016.

Panel A: Prices of Risk

		5% Significance Level			10% Significance Level		
	N	$\lambda > 0$	Average λ	Dispersion λ	$\lambda > 0$	Average λ	Dispersion λ
National Housing Risk	135	20	0.16%	0.13%	27	0.14%	0.12%
Local Housing Risk	135	44	0.13%	0.09%	48	0.13%	0.09%
Zip Code-Level Housing Residual Risk	135	31	0.12	0.06	43	0.10	0.06

Panel B: 5% Significance Level

	λ^{MSA} and λ^{US}	λ^{MSA} and λ^{ZIP}	λ^{US} and λ^{ZIP}	All three
Number of MSAs	6	15	5	3
Share of Population	2.43%	11.90%	10.91%	1.29%
Share of Zip Codes	3.50%	14.49%	10.01%	1.69%
		All Risks		
	>= 1	>= 2	= 3	
Number of MSAs	72	20	3	
Share of Population	62.59%	22.66%	1.29%	
Share of Zip Codes	61.14%	24.61%	1.69%	

Panel C: 10% Significance Level

	λ^{MSA} and λ^{US}	λ^{MSA} and λ^{ZIP}	λ^{US} and λ^{ZIP}	All three
Number of MSAs	8	21	6	4
Share of Population	4.10%	18.24%	12.04%	2.41%
Share of Zip Codes	6.19%	18.76%	12.19%	3.87%
		All Risks		
	>= 1	>= 2	= 3	
Number of MSAs	87	27	4	
Share of Population	77.51%	29.55%	2.41%	
Share of Zip Codes	73.92%	29.40%	3.87%	

Table 4: Housing Risk Premia and Underdiversification

This table reports marginal effects from probit model estimations in which the dependent variables are indicators for whether zip-code level housing residual risk (λ^{ZIP}) (columns (1)-(2)), local housing risk (λ^{MSA}) (columns (3)-(4)), and national housing risk (λ^{US}) (columns (5)-(7)), are positive and statistically significant at the 10% level or higher, respectively. High–Homeownership is an indicator for MSAs with homeownership above the sample 75th percentile. Local–Landlord is an indicator for MSAs with the number of local landlord rental housing units as a percentage of total housing units above the sample 75th percentile. National–Landlord is an indicator for MSAs with the number of national landlord rental housing units as a percentage of total housing units above the sample 75th percentile. Second–Homes is an indicator for MSAs with second homes units as a percentage of total housing units above the sample 75th percentile. Additional control variables include the price to income ratio. Refer to Table A.1 of the main text Appendix for detailed variable definitions. Robust standard errors are reported in parentheses. Coefficients that are significant at a 10%, 5%, and 1% significance level are marked with *,**,***, respectively.

	Dependent Variable I($\lambda > 0$)						
	λ^{ZIP}		λ^{MSA}		λ^{US}		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High–Homeownership	0.218**	0.284***				0.024	
	(0.101)	(0.106)				(0.084)	
Local–Landlord			0.283***	0.285**			-0.001
			(0.100)	(0.120)			(0.101)
National–Landlord		-0.042		0.112	0.203**	0.207**	0.206**
		(0.095)		(0.099)	(0.090)	(0.091)	(0.091)
Second–Homes		-0.188**		-0.097		0.081	0.085
		(0.095)		(0.105)		(0.087)	(0.092)
Price to Income Ratio		0.070		-0.026		-0.005	-0.007
		(0.049)		(0.058)		(0.042)	(0.047)
Nobs	125	125	125	125	134	125	125
Pseudo R^2	0.031	0.067	0.050	0.064	0.044	0.059	0.058

Table 5: ZTRAX Zip Code-Level House Value Indexes

This table reports descriptive statistics of estimated prices of risk that are positive and statistically significant, based on single–sided tests. We construct our own zip code–level indexes using ZTRAX data. National and local housing risk indexes are from Zillow ZHVI data. We estimate prices of risk for the national housing risk (λ^{US}), local housing risk (λ^{MSA}) and zip code-level housing residual risk (λ^{ZIP}) for each MSA with at least 20 zip codes. In Panel A, we report the number of estimated MSAs (N) and the number of MSAs where each price of risk is positive and significant ($\lambda > 0$). Further, we report average λ and the cross–sectional dispersion in λ (Dispersion) within the MSAs where each respective price of risk is positive and significant. In Panel B, we look at how many MSAs have one or more significant prices of risk. We also report the share of the full sample population (Share of Population) and number of zip codes (Share of Zip Codes) these MSAs represent. Significance is at the 10% level single–sided using Newey and West (1994) standard errors with 4 lags. Panel C reports marginal effects from probit model estimations in which the dependent variables are indicators for whether λ^{ZIP}, λ^{MSA}, and λ^{US}, are positive and statistically significant at the 10% level or higher, respectively. High–Homeownership is an indicator for MSAs with homeownership above the sample 75th percentile. Local–Landlord is an indicator for MSAs with the number of local landlord rental housing units as a percentage of total housing units above the sample 75th percentile. National–Landlord is an indicator for MSAs with the number of national landlord rental housing units as a percentage of total housing units above the sample 75th percentile. Additional control variables include the price to income ratio and second–Homes which is an indicator for MSAs with second homes units as a percentage of total housing units above the sample 75th percentile. Refer to Table A.1 of the main text Appendix for detailed variable definitions. The sample period is April 1996 – December 2016.

Panel A: Prices of Risk				
	N	$\lambda > 0$	Average λ	Dispersion λ
National Housing Risk	99	31	0.13%	0.08%
Local Housing Risk	99	28	0.12%	0.07%
Zip Code-Level Housing Residual Risk	99	47	0.05	0.03

Panel B: All Risks			
	>= 1	>= 2	= 3
Number of MSAs	75	26	5
Share of Population	87.37%	42.57%	15.10%
Share of Zip Codes	82.21%	35.89%	14.50%

Panel C: Probit Marginal Effects				
	Dependent Variable I($\lambda > 0$)			
	λ^{ZIP}	λ^{MSA}	λ^{US}	
	(1)	(2)	(3)	(4)
High–Homeownership	0.250**		-0.001	
	(0.117)		(0.116)	
Local–Landlord		0.280**		-0.063
		(0.141)		(0.139)
National–Landlord	0.056	0.040	0.243**	0.244**
	(0.117)	(0.107)	(0.108)	(0.107)
Controls	Yes	Yes	Yes	Yes
Nobs	93	93	93	93
Pseudo R^2	0.036	0.187	0.064	0.066

Table 6: Estimation using Unsmoothed Housing Returns

This table reports descriptive statistics of estimated prices of risk that are positive and statistically significant, based on single–sided tests. We estimate prices of risk for the national housing risk (λ^{US}), local housing risk (λ^{MSA}) and zip code-level housing residual risk (λ^{ZIP}) for each MSA with at least 20 zip codes. The multi–factor model is estimated using unsmoothed housing excess returns based on an AR(3) specification. In Panel A, we report the number of estimated MSAs (N) and the number of MSAs where each price of risk is positive and significant ($\lambda > 0$). We also report the percentage of MSAs in common with our main results (Overlap). Further, we report average λ and the cross–sectional dispersion in λ (Dispersion) within the MSAs where each respective price of risk is positive and significant. In Panel B, we look at how many MSAs have one or more significant prices of risk. We also report the share of the full sample population (Share of Population) and number of zip codes (Share of Zip Codes) these MSAs represent. Significance is at the 10% level single–sided using Newey and West (1994) standard errors with 4 lags. Panel C reports marginal effects from probit model estimations in which the dependent variables are indicators for whether λ^{ZIP}, λ^{MSA}, and λ^{US}, are positive and statistically significant at the 10% level or higher, respectively. High–Homeownership is an indicator for MSAs with homeownership above the sample 75th percentile. Local–Landlord is an indicator for MSAs with the number of local landlord rental housing units as a percentage of total housing units above the sample 75th percentile. National–Landlord is an indicator for MSAs with the number of national landlord rental housing units as a percentage of total housing units above the sample 75th percentile. Additional control variables include the price to income ratio and second–Homes which is an indicator for MSAs with second homes units as a percentage of total housing units above the sample 75th percentile. Refer to Table A.1 of the main text Appendix for detailed variable definitions. The sample period is April 1996 – December 2016.

Panel A: Prices of Risk

	N	$\lambda > 0$	Average λ	Dispersion λ
National Housing Risk	135	26	0.05%	0.05%
Local Housing Risk	135	52	0.08%	0.06%
Zip Code-Level Housing Residual Risk	135	71	0.14	0.09

Panel B: All Risks

	>= 1	>= 2	== 3
Number of MSAs	105	41	3
Share of Population	90.09%	50.20%	1.17%
Share of Zip Codes	87.55%	47.62%	1.21%

Panel C: Probit Marginal Effects

	Dependent Variable I($\lambda > 0$)			
	λ^{ZIP}	λ^{MSA}	λ^{US}	
	(1)	(2)	(3)	(4)
High–Homeownership	0.214**		0.086	
	(0.100)		(0.088)	
Local–Landlord		0.261**		0.074
		(0.121)		(0.119)
National–Landlord	0.016	0.077	0.190**	0.191**
	(0.103)	(0.100)	(0.090)	(0.091)
Controls	Yes	Yes	Yes	Yes
Nobs	125	125	125	125
Pseudo R^2	0.050	0.071	0.054	0.046

Table A.1: Variable Definitions

This table provides the definitions of the main variables used in this paper.

Main Variables:	**Definition:**
U.S. Stock Market Risk	Proxied by U.S. stock market excess returns. U.S. stock market returns and risk-free rate are obtained from the online data library of Kenneth R. French. Data are monthly from 1996-2016.
National Housing Risk	Proxied by U.S. housing market excess returns. Data on U.S. housing returns are from Zillow's ZHVI or from Zillow's ZTRAX database, respectively. Data are monthly from 1996-2016.
Local Housing Risk	Proxied by local MSA excess returns. Data on MSA-level housing returns are from Zillow's ZHVI or from Zillow's ZTRAX database, respectively. Data are monthly from 1996-2016.
Zip Code-Level Housing Residual Risk	Proxied by idiosyncratic volatility which is estimated as the standard deviation of residuals of a time–series regression of zip code excess returns on the U.S. stock market, U.S. housing market and local MSA excess returns. Data on zip code-level housing returns are from Zillow's ZHVI or from Zillow's ZTRAX database, respectively. Data are monthly from 1996-2016.
High–Homeownership	Indicator equal to 1 for MSAs with homeownership above the sample 75th percentile. To obtain MSA homeownership, we look at the percentage of households that are owner-occupiers relative to all households within an MSA. Data on homeownership are from IPUMS (item OWNERSHP). Yearly sample, 2000 (decennial U.S. Census 1% sample) and 2005-2016 (ACS 1% sample).
Local–Landlord	Indicator equal to 1 for MSAs with number of local landlord housing units over total housing units above the sample 75th percentile. Local landlord housing units are obtained by first multiplying total housing units by 1-homeownership to obtain total rental housing units. Next, we subtract national landlord housing units to obtain local landlord housing units. We then obtain data on second homes housing units. Finally, we subtract second homes from local landlord housing units to obtain our final measure of local landlord housing units. Data on total housing units are from the U.S. Census Bureau. Data on homeownership are from IPUMS (item OWNERSHP). Data on second homes units are from IPUMS NHGIS (item Vacancy Status). Data on national landlord units are from S&P Global Market Intelligence SNL Real Estate Property. Yearly sample, 2000 (decennial U.S. Census 1% sample) and 2005-2016 (ACS 1% sample).
National–Landlord	Indicator equal to 1 for MSAs with number of national landlord housing units over total housing units above the sample 75th percentile. Data on total housing units are from the U.S. Census Bureau. Data on real estate companies property holdings are from S&P Global Market Intelligence SNL Real Estate Property. Yearly data from 1996-2016.

(Table continues on next page.)

Main Variables:	Definition:
Second–Homes	Indicator equal to 1 for MSAs with second homes units over total housing units above the sample 75th percentile. Data on second homes units are from IPUMS NHGIS (item Vacancy Status). Data on total housing units are from the U.S. Census Bureau. Data are available for 2000 (decennial U.S. Census 1% sample), 2007-2011 (ACS 5-year year summary file), and 2012-2016 (ACS 5-year year summary file).
Price to Income Ratio	Median price divided by median income. We obtain MSA montly price data from Zillow, and we compute the median across months. To obtain MSA income we compute the median total money income of all household members age 15+ within each MSA. Data on MSA-level housing prices is from Zillow's ZHVI, monthly from 1996-2016. Data on household income are from IPUMS (item HHINCOME) for 2000 (decennial U.S. Census 1% sample) and 2005-2016 (ACS 1% sample).
Adjusted High–Homeownership	Indicator equal to 1 for MSAs with adjusted homeownership above the sample 75th percentile. We adjust MSA homeownership by adding the percentage of small rental units (1-4 units) owned by individual investors. The data on small rental units owned by individual investors for the U.S. is from the RHFS (Rental Housing Finance Survey) in 2015 and 2018. We omit the 2012 RHFS as it does not contain data on 1-unit rentals which are a large part of the rental market. Using the RHFS items OWNENT, NUMUNITS_R and WEIGHT, we calculate the number of small rental units owned by individual investors across the U.S. Next, for each MSA, we assign a number of small rental units owned by individual investors proportional to their population relative to the U.S. population. Finally, we divide small rental units by total units to generate our adjustment. Yearly sample, 2000 (decennial U.S. Census 1% sample) and 2005-2016 (ACS 1% sample).
Adjusted Local–Landlord	Indicator equal to 1 for MSAs with adjusted number of local landlord housing units over total housing units above the sample 75th percentile. We adjust the number of local landlord housing units by subtracting the number of small rental units (1-4 units) owned by individual investors. The data on small rental units owned by individual investors for the U.S. is from the RHFS (Rental Housing Finance Survey) in 2015 and 2018. We omit the 2012 RHFS as it does not contain data on 1-unit rentals which are a large part of the rental market. Using the RHFS items OWNENT, NUMUNITS_R and WEIGHT, we calculate the number of small rental units owned by individual investors across the U.S. Next, for each MSA, we assign a number of small rental units owned by individual investors proportional to their population relative to the U.S. population. Finally, we subtract these small rental units from the local landlord units for each MSA. Yearly sample, 2000 (decennial U.S. Census 1% sample) and 2005-2016 (ACS 1% sample).
Number of Zip Codes	Average number of zip codes per MSA in hundreds from 1996–2016. Data on zip code-level housing returns is from Zillow's ZHVI. Data are monthly from 1996-2016.

Figure 1: Histogram of Risk Exposure t–statistics

In this figure, we plot the distribution of Newey and West (1994) adjusted t–statistics for the estimated risk exposures of zip code–level housing excess returns on the U.S. stock market risk factor, national housing risk factor and local housing risk factor. The sample period is April 1996 – December 2016.

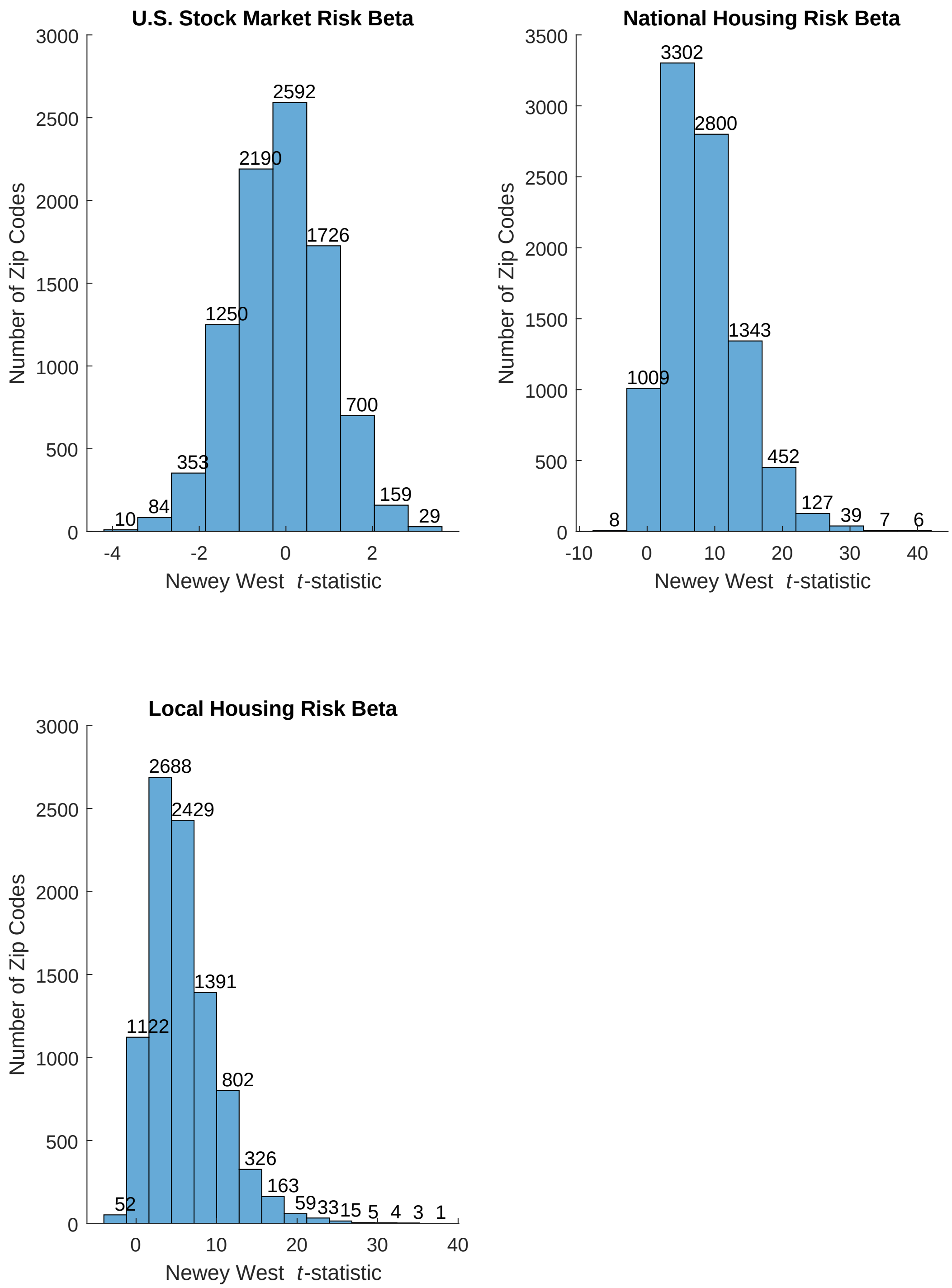
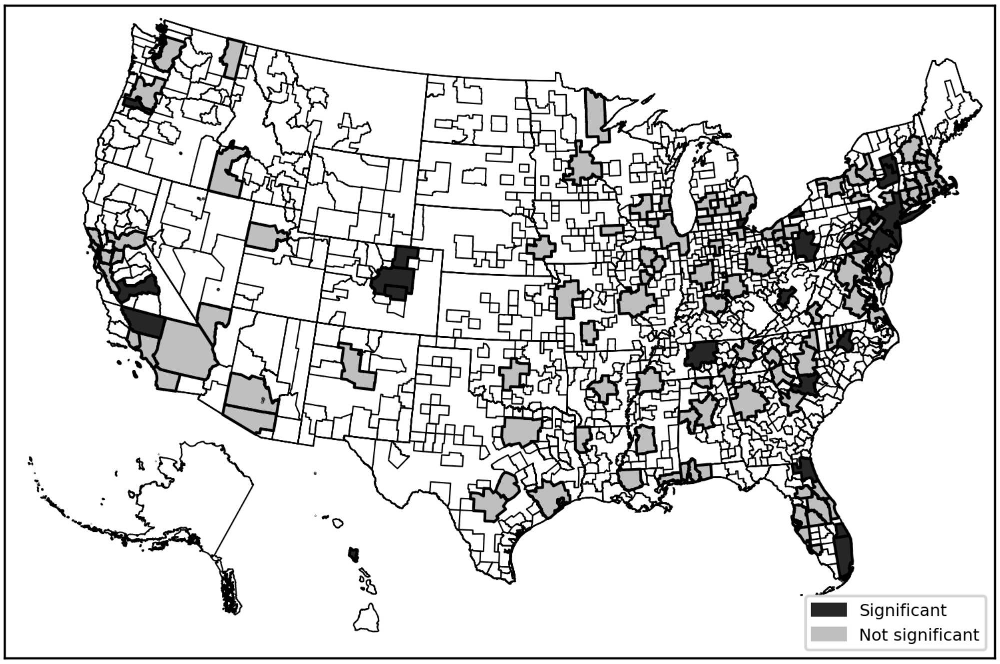


Figure 2: Prices of Risk Estimates across the U.S.

In this figure, we plot geographical heat maps of the U.S. containing the 135 MSAs for which we estimate prices of risk. In each panel, we indicate in black MSAs that have a statistically positive significant price of risk, and in gray MSAs where it is statistically insignificant. In Panel (a), we plot a heat map for the price of the national housing risk; in Panel (b) for the local housing risk; in Panel (c) for zip code-leve housing residual risk; and in Panel (d) for the U.S. stock market risk. The sample period is April 1996 – December 2016.

(a) National Housing Risk



(b) Local Housing Risk

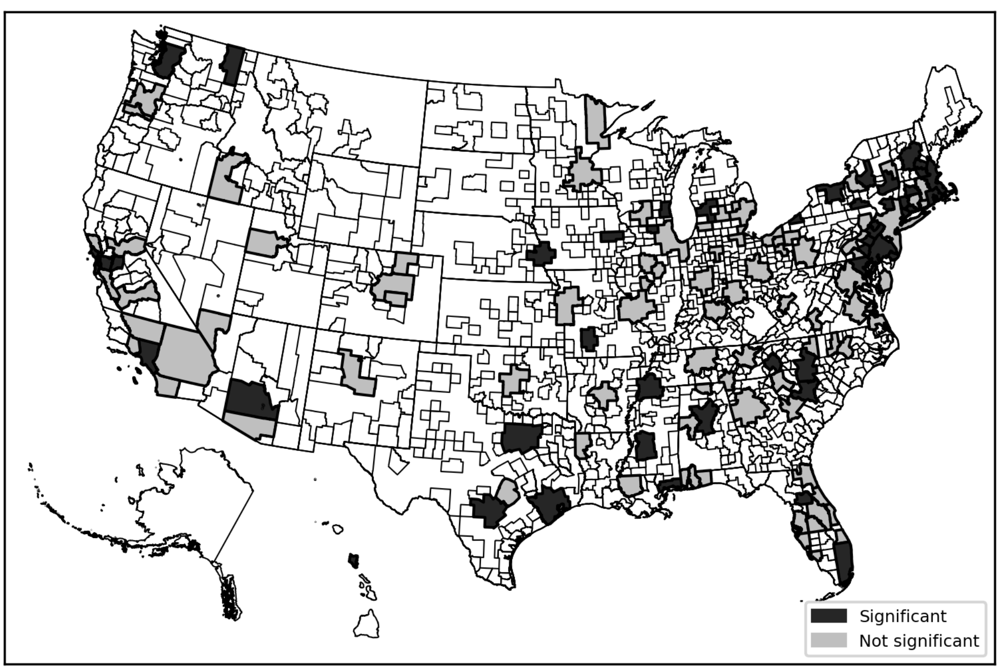
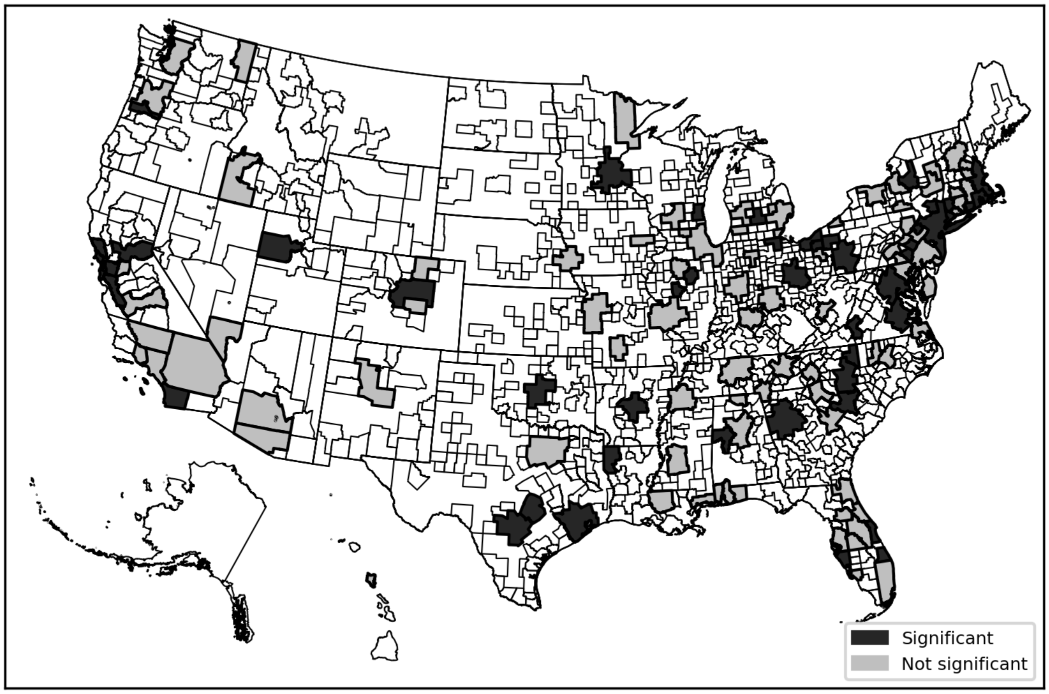


Figure 2: Prices of Risk Estimates across the U.S. (cont.)

(c) Zip Code-Level Housing Residual Risk



(d) U.S. Stock Market Risk