

NBER WORKING PAPER SERIES

CAN OWNING A HOME HEDGE THE RISK OF MOVING?

Todd M. Sinai  
Nicholas S. Souleles

Working Paper 15462  
<http://www.nber.org/papers/w15462>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
October 2009

We are grateful to Dimo Primatov for helpful research assistance and seminar participants at the National Bureau of Economic Research, Stanford University, the University of Amsterdam, and Wharton for useful feedback. Sinai thanks the Research Sponsors Program of the Zell-Lurie Real Estate Center at Wharton for financial support. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2009 by Todd M. Sinai and Nicholas S. Souleles. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Can Owning a Home Hedge the Risk of Moving?

Todd M. Sinai and Nicholas S. Souleles

NBER Working Paper No. 15462

October 2009

JEL No. E21,G11,G12,J61,R21,R23,R31

**ABSTRACT**

Conventional wisdom holds that one of the riskiest aspects of owning a house is the uncertainty surrounding its sale price, especially if one moves to another housing market. However, households who sell a house typically buy another house, whose purchase price is also uncertain. We show that for such households, home owning often hedges their net exposure to housing market risk, because their sale price covaries positively with house prices in their likely new market. That expected covariance is much higher than previously recognized because there is considerable heterogeneity across city pairs in how much house prices covary and households tend to move between the highly correlated housing markets. Taking these two considerations into account increases the estimated median expected correlation in real house price growth across MSAs from 0.35 to 0.60. Moreover, we show that households' decisions whether to own or rent are sensitive to this "moving-hedge" value. We find that the likelihood of home owning for a mobile household is more than one percentage point higher when the expected house price covariance rises by 38 percent (one standard deviation). This effect attenuates as a household's probability of moving diminishes and thus the moving-hedge value declines.

Todd M. Sinai

University of Pennsylvania, Wharton School

1465 Steinberg Hall - Dietrich Hall

3620 Locust Walk

Philadelphia, PA 19104-6302

and NBER

sinai@wharton.upenn.edu

Nicholas S. Souleles

Finance Department

The Wharton School

2300 SH-DH

University of Pennsylvania

Philadelphia, PA 19104-6367

and NBER

souleles@wharton.upenn.edu

Conventional wisdom holds that one of the riskiest aspects of owning a house is the uncertainty surrounding its sale price, especially if one moves to another housing market. It is now well appreciated that house prices can be quite volatile. Between the end of 2005 and the end of 2008, real house prices fell by more than 31 percent, according to the Case-Shiller 10-city composite house price index. Over the prior five years, real house prices in the same cities rose by almost 73 percent. Similarly, after real house prices rose substantially during the 1980s, they fell by 26 percent between 1990 and 1997.

Historically, analysts have concluded that this volatility in house prices makes home owning risky. Since the primary residence comprises about two-thirds of the median homeowner's assets (2004 Survey of Consumer Finances), a gain or a loss on a house could have a sizeable effect on a household's balance sheet. In addition, nearly 45 percent of households move within a five-year period and one-fifth of such households leave their metropolitan area. A loss on a house could impair their ability to purchase their next house. These are some of the reasons Case *et al.* (1993) argue for using house price derivatives to help households offset house price volatility. Similarly, in some cities home equity insurance products have been created, enabling households to guarantee (for a fee) that their house values will not fall below some threshold. [Caplin *et al.* (2003)]

In this paper, we show that for many households home owning is not as risky as conventionally assumed. Households who expect to sell their house will still have to live somewhere afterwards and the cost of that subsequent housing is uncertain. For many home owners, the sale price of their current house positively commoves with the purchase price of their next house, reducing the expected volatility in the net cost of selling one house and buying another. In effect, owning a house provides a hedge against the uncertain purchase price of a

future house. By contrast, a household who rents avoids the uncertainty from selling its home but leaves itself unhedged against the volatility of future housing costs after moving.

A few recent papers have recognized that it is the sale price *net* of the subsequent purchase price, rather than the sale price alone, that matters for housing risk. [Ortalo-Magne and Rady (2002), Sinai and Souleles (2005), Han (2008)] However, these papers – along with conventional wisdom – assume that the correlation in house prices across housing markets is low, which implies that owning a house provides a poor hedge against future housing costs for households who face some chance of moving to a different market. The literature has instead emphasized that home owning can be a good hedge against buying a larger house in the *same* market.<sup>1</sup>

While it is correct that house prices do not covary much when one considers the U.S. as a whole, that unconditional average masks two important factors. First, there is considerable heterogeneity across city pairs within the U.S. in how much their house prices covary, ranging from negative covariances to very highly positive covariances. Second, households do not move to random locations; instead, they tend to move between highly covarying housing markets. Our first contribution is to show that because of these two considerations, for many households the expected covariance in house prices – where the expectation is weighted by the household's own probabilities of moving to each other location – is quite high. For example, the simple unweighted median correlation in house price growth across U.S. metropolitan statistical areas

---

<sup>1</sup> The benefit of home owning as a hedge against future house price risk in other cities is generally undeveloped in prior research. Ortalo-Magne and Rady (2002) illustrate in a simple theoretical model that house prices in one period hedge prices in the next period if the prices covary across the periods, but provide no empirical evidence on the magnitude or effect of the hedge. Sinai and Souleles (2005) show theoretically how sale price risk depends on the covariance between house prices in the current and future housing markets, but their primary empirical focus is on how home owning hedges against volatility in housing costs within a given housing market. Han (2008) distinguishes within- and out-of-state moves in a structural model of housing consumption using the Panel Study of Income Dynamics. However, she does not know where households move if they move out of state, so she does not estimate cross-state covariances. Cocco (2000) considers hedging the cost of trading up to a larger house within the same housing market, by liquidity constrained households.

(MSAs) is 0.35. When we account for where households are likely to move, the baseline effective correlation faced by the median household rises substantially, to 0.60. The 75<sup>th</sup> percentile household enjoys an even higher correlation, of 0.89.

Because households' effective expected covariances are quite high, owning a house can provide a valuable hedge against house price risk, especially for households who are likely to move. This includes households who do not know exactly where they are going to move. As long as they are likely to move to positively covarying markets, home owning helps hedge their expected purchase price risk.<sup>2</sup>

Our second contribution is to show that households' tenure decisions (to rent versus own) appear to be sensitive to this "moving-hedge" benefit of owning. We bring three sources of variation to bear on this issue. First, the typical household *across* different MSAs may have a bigger, smaller, or even negative hedging benefit of home owning depending on the variance of house prices in the local MSA and their covariances with prices in other MSAs. Second, *within* an MSA, households differ in their expected covariances because they differ in their likelihoods of moving to each of the other MSAs. Third, the effect of the expected house price covariance should be attenuated for households who, for demographic reasons, appear unlikely to move in the near future. Households who do not anticipate moving have little need for a hedge against future house prices, and thus differences in the potential moving-hedge benefit of owning should have little effect on their tenure decisions.

---

<sup>2</sup> In order to highlight how the risk of home owning is affected by the house's moving-hedge properties, we abstract from the effect of housing finance on household risk. Financing any volatile asset with high leverage creates risk, a topic which has been considered at length in prior research. For example, for leveraged, liquidity-constrained households, downturns in prices can lead to lock-in and more volatile consumption and prices. [Chan (2001), Stein (1995), Genesove and Mayer (1997, 2001), Hurst and Stafford (2004), Lustig and Van Nieuwerburgh (2005), Li and Yao (2007)]

We use household-level data on homeownership and moving probabilities, and MSA-level estimates of house price variances and covariances, to identify the effect of expected house price covariances on homeownership decisions. We use demographic characteristics such as age, marital status, and occupation to impute the probability of moving for each household. We impute the odds of a household moving to specific MSAs, conditional on moving at all, by applying the actual geographic distribution of moves by other households in similar industry or age categories in the originating MSA. This combination of MSA and household level variation enables us to identify the effect of expected house price covariances on homeownership decisions while controlling separately for MSA and household characteristics.

Overall, we find that for a household with an average expected length-of-stay in the house, the likelihood of home owning increases by 0.6 percentage points when the expected covariance rises by one standard deviation. This effect is larger for households with a higher likelihood of moving to another MSA. A household who is imputed to have a one-in-three chance of moving would have a 1 to 2 percentage point higher likelihood of home owning if faced with a one standard deviation increase in expected covariance. The results are robust to different ways of imputing the likelihood of a household's moving to the various MSAs, such as instrumenting for the actual moving patterns of households with the patterns we would predict if households moved based on the distribution of their industry's employment or their occupations across MSAs.

In the next section, we present a simple theoretical framework to illustrate the moving-hedge benefit of owning and to motivate our empirical tests. In Section 2, we estimate households' effective covariances between house price growth in their current markets and in their expected future markets, and explain why conventional wisdom has assumed those

covariances are low when they are actually quite high. Section 3 describes the various data sources we use. The empirical identification strategy and results are presented in Section 4. Section 5 briefly concludes.

### *1. Theoretical framework*

In this section, we illustrate how owning a house can hedge against the house price risk from future moves. This illustration will also provide guidance for the empirical tests that follow. Our exposition generally follows Sinai and Souleles (2005). To simplify, and focus attention on the moving-hedge benefit of owning, we abstract from some other important issues, such as leverage and down payment requirements, taxes, and moving costs, which would operate in addition to the hedging benefit. (Such issues will be taken into account in our empirical work.)

Since our focus is on how owning a house in one city can hedge against house price volatility in the next city, we will consider a representative household who initially lives in some city,  $A$ , and then moves to another city,  $B$ . To simplify, we assume that the household knows with certainty that it will live in  $A$ , and then in  $B$ , for  $N$  years each, after which it will die. (In our empirical work, we will recognize that  $N$  can vary across households, with some expecting to not move very often and others expecting to move more frequently, and that there are multiple destination cities to which households could move.) At birth, labeled year 0, the household chooses whether to be a homeowner in both locations or a renter in both locations.<sup>3</sup> The household chooses its tenure mode to maximize its expected utility of wealth net of total housing costs, or equivalently to minimize its total risk-adjusted housing costs.

---

<sup>3</sup> The desired quantity of housing services is normalized to be one unit in each location. For convenience, rental units and owner-occupied units, in fixed supply and together equal to the number of households, both provide one unit of housing services. The results below can be generalized to allow the services from an owner-occupied house to exceed those from renting, perhaps due to agency problems. Additional extensions are discussed in Sinai and Souleles (2005).

The cost of obtaining a year's worth of housing services is the rent, denoted by  $\tilde{r}_t^A$  in city  $A$  in year  $t$ , and  $\tilde{r}_t^B$  in city  $B$ . The tildes denote that the rent in year  $t$  is not known at time zero, because rents fluctuate due to shocks to underlying housing demand and supply. To allow for correlation in rents (and, endogenously, in house prices) across cities, we assume that rents in the two locations follow correlated AR(1) processes:  $r_n^A = \mu^A + \varphi r_{n-1}^A + k(\eta_n^A + \rho \eta_n^B)$  and  $r_n^B = \mu^B + \varphi r_{n-1}^B + k(\rho \eta_n^A + \eta_n^B)$ , where  $\varphi \in [0, 1]$  measures the persistence of rents,  $\mu^A$  and  $\mu^B$  measure the expected level or growth rate of rents (depending on  $\varphi$ ), and the shocks  $\eta^A$  and  $\eta^B$  are independently distributed  $\text{IID}(0, \sigma_A^2)$  and  $\text{IID}(0, \sigma_B^2)$ .  $\rho$  parameterizes the spatial correlation in rents (and in house prices) across the two locations, with  $\rho=0$  implying independence and  $\rho=1$  implying perfect correlation. To control the total magnitude of housing shocks incurred as  $\rho$  varies, the scaling constant  $\kappa$  can be set to  $1/(1+\rho^2)^{1/2}$ . For simplicity in this exposition, we will set the persistence term  $\varphi$  to 0. We find similar qualitative results with the more realistic assumption of  $\varphi > 0$ . [Case and Shiller (1989)]

From a homeowner's perspective, the lifetime *ex post* cost of owning, discounted to year 0, is  $C_O \equiv P_0^A + \delta^N (\tilde{P}_N^B - \tilde{P}_N^A) - \delta^{2N} \tilde{P}_{2N}^B$ . The  $P_0^A$  term is the initial purchase price in city  $A$ , which is known at time 0. In the last term,  $\tilde{P}_{2N}^B$  is the uncertain residual value of the house in  $B$  at the time of death. It is discounted since death occurs  $2N$  years in the future. Our emphasis in this paper is on the middle term,  $\delta^N (\tilde{P}_N^B - \tilde{P}_N^A)$ , which is the difference between the sale price of the house in  $A$  at time  $N$  and the purchase price of the house in  $B$  at time  $N$ .

For renters, the *ex post* cost of renting is the present value of the annual rents paid:

$$C_R \equiv r_0^A + \sum_{n=1}^{N-1} \delta^n \tilde{r}_n^A + \sum_{n=N}^{2N-1} \delta^n \tilde{r}_n^B.$$



Sinai and Souleles (2005) derive house prices in this setting assuming they endogenously adjust to leave households indifferent between owning and renting. The resulting price in city  $A$ ,  $P_0^A$ , can be expressed as the expected present value of future rents,  $PV(r_0^A)$ , plus the total risk premium the household is willing to pay to own rather than to rent:

$$P_0^A = PV(r_0^A) + \frac{(\pi_R - \pi_O)}{1 - \delta^N} - \frac{\delta^N (\pi_R^B - \pi_O^B)}{1 - \delta^N}. \quad (1)$$

In the second term, the risk premium for owning,  $\pi_O$ , measures the risk associated with the cost  $C_O$  of owning, which in equilibrium reduces the price  $P_0^A$  in equation (1), *ceteris paribus*.

The risk premium for renting,  $\pi_R$ , measures the risk associated with the cost  $C_R$  of renting. Since owning avoids this risk,  $\pi_R$  increases  $P_0^A$ , *ceteris paribus*.  $P_0^A$  also capitalizes  $\Delta\pi^B \equiv (\pi_R^B - \pi_O^B)$ , a net risk premium for renting versus owning in  $B$  that in equilibrium  $P_0^A$  inherits from house prices in  $B$ .

For owners,  $\pi_O$  measures the total house price risk from the three future housing transactions; i.e., the sale of the first house in  $A$ , and the purchase price and subsequent sale price of the second house in  $B$ :

$$\pi_O \equiv \frac{\alpha}{2} \left[ \delta^{2N} \kappa^2 (1 - \rho)^2 (s_A^2 + s_B^2) + \delta^{4N} (s_B^2) \right], \quad (2)$$

where  $s_A^2 \equiv \text{var}(r_A)$  and  $s_B^2 \equiv \text{var}(r_B)$  are the variance of rents in cities  $A$  and  $B$ , respectively, and  $\alpha$  is household risk aversion. Since house prices are endogenously related to rents (as in eq. (1)),

house price volatility follows from rent volatility. Thus we can use  $s_A^2$  and  $s_B^2$  to measure the underlying housing market volatility.<sup>4</sup>

In the final term in eq. (2),  $s_B^2$  reflects the risk associated with the sale of the house in  $B$ , discounted by  $\delta^{(2N)^2}$  since it takes place in year  $2N$ .

The first term in eq. (2) reflects the net risk from the sell-in- $A$  and buy-in- $B$  transactions in year  $N$ , i.e. the risk associated with the *difference* between the purchase price and sale proceeds,  $(\tilde{P}_N^B - \tilde{P}_N^A)$ . The net risk depends on the correlation  $\rho$  between house prices in  $A$  and  $B$ .

The part of the first term inside the brackets can be written as  $\delta^{2N} f(\rho)(s_A^2 + s_B^2)$ , with

$f(\rho) \equiv \kappa^2(1 - \rho)^2 = \frac{(1 - \rho)^2}{1 + \rho^2}$ . If prices in the two markets are uncorrelated, with  $\rho=0$ , then

$f(\rho) = 1$ , so the net risk from the two transactions is simply the sum of the risks of the individual transactions  $(s_A^2 + s_B^2)$ , appropriately discounted. But as the two markets become increasingly correlated, the net risk declines. That is, owning the house in  $A$  helps to hedge against the uncertainty of the purchase in  $B$ . In the polar case when  $\rho=1$ , then  $f(\rho)=0$ , and so the sale and subsequent purchase are expected to fully wash each other out. By contrast, if the prices in the two markets are perfectly negatively correlated, with  $\rho=-1$ , then  $f(\rho)=2$ , and the net risk is twice as large as the sum of the individual risks.

Since  $f(\rho)$  is monotonically decreasing in  $\rho$ , in our empirical work it will be useful to use the approximation

---

<sup>4</sup> If  $\rho > 0$ , the variance of observed rents in a given city includes the contribution of the underlying housing market shocks  $\eta$  from the other city as well:  $s_A^2 = \kappa^2(\sigma_A^2 + \rho^2\sigma_B^2)$  and  $s_B^2 = \kappa^2(\sigma_B^2 + \rho^2\sigma_A^2)$ . In the symmetric case with  $\sigma_A^2 = \sigma_B^2 = \sigma^2$ , then  $s_A^2 = s_B^2 = \sigma^2$ , independent of  $\rho$ . With  $\rho=0$  in eq. (2), the price risks come from the contemporaneous rent shocks:  $\eta_N^A$  on  $P_N^A$ ,  $\eta_N^B$  on  $P_N^B$ , and  $\eta_{2N}^B$  on  $P_{2N}^B$ . With  $\rho > 0$ , the prices would also include the persistent effect of the preceding rent shocks. [See Sinai and Souleles (2005).]

$$\pi_O \approx \frac{\alpha}{2} \left[ -\delta^{2N} \text{cov}(A, B) + \delta^{4N} (s_B^2) \right], \quad (3)$$

where  $\text{cov}(A, B)$  is the covariance of rents (and prices) in  $A$  and  $B$ . The risk premium from owning should decline with this covariance.

For renters, uncertainty comes from not having locked-in the future price of housing services, so the risk of renting is proportional to the discounted sum of the corresponding rent shocks:

$$\pi_R \equiv \frac{\alpha}{2} \left( s_A^2 \sum_{n=1}^{N-1} \delta^{2n} + s_B^2 \sum_{n=N}^{2N-1} \delta^{2n} \right). \quad (4)$$

Turning to the remaining terms in eq. (1), the net risk premium in  $B$  that is capitalized into  $P_0^A$  analogously depends on the net risk of owning versus renting while living in  $B$ :

$$(\pi_R^B - \pi_O^B) \approx \frac{\alpha}{2} s_B^2 \left( \sum_{n=1}^{N-1} (\delta^n)^2 - (\delta^N)^2 \right). \quad (5)$$

Finally, the present value of expected rents in  $A$  increases with the trend  $\mu^A$ :

$$PV(r_0^A) \equiv \left( r_0^A + \mu^A \frac{\delta}{1-\delta} \right) \quad (6)$$

Generalizing this framework to a setting with heterogeneous households,  $P_0^A$  can be thought of as reflecting a household's latent demand for owning. If the household's willingness-to-pay is above the market-clearing house price determined by the marginal homebuyer, the household would own, otherwise it would rent. In our empirical work, we observe the own/rent decision but not latent demand. We will control nonparametrically for differences across MSAs in the clearing prices for houses, by including MSA  $\times$  year fixed effects, and will allow for heterogeneity across households (in particular in  $N$  and  $\rho$ ). This approach allows us to map our

empirical results on the determinants of the tenure decision to inferences about the latent demand for home ownership.

This framework yields several empirical predictions. First, as the covariance in house prices between cities  $A$  and  $B$  rises, a household living in  $A$  should be more likely to own its house. By equation (3), the risk of owning declines with the covariance, because the house acts as a hedge. Thus the price  $P_0^A$  and the demand for owning should increase with the covariance, *ceteris paribus*.

Second, that hedging value should diminish as the likelihood of moving falls. As a household's expected length-of-stay,  $N$ , in city  $A$  increases, the net price risk in eq. (3) is expected to occur further in the future and thus is discounted more heavily. That is, in making its tenure decision a household who expects to be mobile (small  $N$ ) should be more sensitive to the moving-hedge benefit of owning, since its move will occur sooner and thus the net price risk will be larger in present value. Conversely, a household who expects it will never move has no need to worry about future housing markets. Thus the demand for owning should decline with the interaction of  $N$  and  $cov(A,B)$ .<sup>5</sup>

These cross-market implications operate in addition to the within-market implications already empirically established in Sinai and Souleles (2005). The key implication they tested is that the effect of local rent volatility  $s_A^2$  on demand generally increases with the horizon ( $N$ ).

Households with longer expected horizons are exposed to a larger number of rent shocks (in  $\pi_R$ ),

---

<sup>5</sup> These results generalize to the case when  $\varphi > 0$ . First,  $P_0^A$  still increases with  $\rho$ ,  $\partial P_0^A / \partial \rho > 0$ . This is because  $\partial \pi_0 / \partial \rho$  remains negative, and  $(\pi_R^B - \pi_0^B)$  is independent of  $\rho$  given  $s_B^2$ . When  $\varphi > 0$ ,  $\pi_R$  is no longer independent of  $\rho$ , but  $\partial \pi_R / \partial \rho$  is positive. That is, a higher covariance also increases the amount of rent risk, which reinforces (though for realistic parameters is quantitatively smaller than) the effect of the reduced price risk due to  $\partial \pi_0 / \partial \rho < 0$ . Second,  $\partial P_0^A / \partial \rho$  generally declines with  $N$  (i.e., the interaction term  $\partial^2 P_0^A / \partial \rho \partial N < 0$ ) for realistic parameters and  $N$  not too small ( $N > 3$ ).

whereas the expected sale price risk (in  $\pi_O$ ) comes further in the future and hence is discounted more heavily. Thus the demand for owning should increase with the interaction of  $N$  and  $s_A^2$  – an implication that Sinai and Souleles confirm empirically. Here we focus instead on the cross-market implications described above.

## 2. *The covariance of house prices across MSAs*

The previous section established that the value of owning a house as a hedge against future moving depends on how much house prices covary across housing markets. Most analysts have concluded that there is little covariance in prices across markets because the national average covariance (equally weighting MSAs) is fairly low. However, that simple average masks three important factors that together often cause the effective covariance faced by households to be quite high. First, there is considerable heterogeneity across housing markets in their covariances with other housing markets. The national average covariance obscures this heterogeneity. Second, households do not move at random. Instead, they are more likely to move among more highly covarying housing markets. The framework in Section 1 assumed that households knew with certainty that they would move from city  $A$  to city  $B$ . In practice, there are a number of cities to which a given household might move with some probability. For that household, the value of the moving hedge depends upon its expected covariance, the probability-weighted average of the covariances of house prices in its current market with house prices in each of its other possible subsequent housing markets. Because of the systematic moving, the average household's expected covariance is higher than the average covariance across MSAs (equally- or population-weighted).

Third, the distribution of the expected covariances is skewed, with a longer lower tail. Because of this, the median household's expected covariance is even higher than that of the average household. That is, while many households have high expected covariances, a minority has very low expected covariances, which lowers the average.

The heterogeneity across and within housing markets can be seen in Figure 1, which graphs each MSA's distribution of house price correlations, for a subsample of the largest MSAs.<sup>6</sup> We compute the correlations in real annual growth in the OFHEO constant-quality MSA-level house price index over the 1980 to 2005 time period. The OFHEO index is computed using repeat sales of single-family houses with conforming mortgages. While the index is widely believed to understate effective house price volatility because it fails to take into account differences in liquidity between housing booms and busts, it is available for a long period for many different MSAs, making it the best data set available for our purposes.

In Figure 1, each vertical grey/black bar represents a metropolitan area. The bottom of each bar is set at the fifth percentile of the MSA's house price correlations with each of the other MSAs, where each MSA is equally weighted and the correlation of an MSA with itself (which would equal one) is excluded. Moving up a bar, the bar turns from grey to black at the 25<sup>th</sup> percentile, then from black to grey at the 75<sup>th</sup> percentile, and the grey portion ends at the 95<sup>th</sup> percentile. Thus the black part of the bar covers the interquartile range of correlations across MSAs, and the entire grey/black bar covers the 5<sup>th</sup> percentile to the 95<sup>th</sup> percentile.

The first thing to note in Figure 1 is that there is substantial heterogeneity across MSAs in their house price correlations with the rest of the country, since the bars for some MSAs start and end higher than the others. For example, consider the distribution of correlations  $\{\rho_{ATL,j}\}$

---

<sup>6</sup> The figure uses the correlation rather than the covariance because the former is easier to interpret visually. The conclusions are the same using the covariance.

between house prices in Atlanta (the first city in the figure) and each of the other cities  $j$ . The 5<sup>th</sup> percentile correlation is  $-0.05$ , and the 95<sup>th</sup> percentile correlation is  $0.73$ . By contrast, for Austin the corresponding percentiles are lower,  $-0.40$  and  $0.54$ . San Antonio has the lowest 5<sup>th</sup> and 95<sup>th</sup> percentiles, ranging from  $-0.50$  to  $0.47$ . The highest among the 5<sup>th</sup> percentiles of correlations is in Miami ( $0.13$ ), and the highest of the 95<sup>th</sup> percentiles is in New York ( $0.94$ ).

While the 5<sup>th</sup> through the 95<sup>th</sup> percentile correlations always overlap across cities, some of the interquartile ranges (the black bars) do not overlap. In Atlanta, the interquartile range of the correlations runs from  $0.27$  to  $0.50$ , which does not overlap at all with the interquartile range in Austin, which runs from  $-0.22$  to  $0.17$ .

The second interesting fact apparent in Figure 1 is that the within-MSA heterogeneity in correlations (with other MSAs) also varies considerably across MSAs. This can be seen by the length of the grey and black bars. MSAs whose bars are stretched out relative to the other MSAs have more heterogeneity in their correlations, being relatively uncorrelated with some other MSAs and relatively highly correlated with others. For example, New York has a  $-0.22$  correlation with its 5<sup>th</sup> percentile correlation city, but a  $0.94$  correlation with its 95<sup>th</sup> percentile city. There is relatively less heterogeneity within Minneapolis, where the corresponding correlations range from  $0.12$  (5<sup>th</sup> percentile) to  $0.68$  (95<sup>th</sup> percentile). There is also significant heterogeneity in the sizes of the interquartile ranges. The widest interquartile range is in Salt Lake City, which spans from  $-0.18$  to  $0.33$ . The narrowest range is in Minneapolis. However, the cities with the widest 5<sup>th</sup> to 95<sup>th</sup> percentile ranges are not necessarily the same ones with the widest interquartile ranges. For example, New York's interquartile range runs from  $0.12$  to  $0.53$ , about the same as Nashville, even though Nashville's 5<sup>th</sup> to 95<sup>th</sup> range is much tighter. San

Jose's interquartile range is fairly tight at 0.15 to 0.38, even though the 5<sup>th</sup> to 95<sup>th</sup> percentile range is middle-of-the-road.

In Figure 2, we take into account where households are likely to move, by weighting the correlations by the probability of moving. In this figure, we use data from the U.S. Department of the Treasury's County-to-County Migration Patterns to impute a household's likelihood of moving from a given MSA to each other MSA. The Treasury data uses the addresses listed on tax returns to determine whether a household moved. It aggregates the gross flows across counties and reports, for each county pair, the number of tax returns annually in which the taxpayers moved from the origination county to the destination county. We aggregated the counties into MSAs, and for each MSA computed the fraction of its taxpaying households who move to each of the other MSAs. We use the resulting fractions to measure the probabilities of any household in the MSA making the corresponding transition. These moving-shares are the weights for computing the distributions in Figure 2. Note that the figure considers only out-of-MSA moves.

When we weight by where households typically move, it becomes clear that most households face much higher effective correlations than indicated in the first figure, because households are more likely to move to more highly correlated MSAs. In Figure 2, the distributions of correlations shift upwards in nearly every MSA.<sup>7</sup> In many MSAs the top of the black bar shifts up near the top of the grey bar. This implies that the entire top quartile of correlations for these MSAs is close to one. (It is impossible to obtain a correlation of exactly

---

<sup>7</sup> The exceptions are minor and tend to occur in cities where there are relatively few high-correlation destinations. Every reported percentile cutoff increases once we weight by the moving probabilities, except: Chicago's 75<sup>th</sup> and 95<sup>th</sup> percentiles decline from 0.70 to 0.67 and 0.81 to 0.75, respectively; Indianapolis' 95<sup>th</sup> percentile declines from 0.80 to 0.79; Phoenix's 75<sup>th</sup> percentile declines from 0.49 to 0.47; and Salt Lake City's 95<sup>th</sup> percentile declines from 0.55 to 0.51. Even in those cities, the rest of the distribution shifts upward considerably. For example, Chicago's median correlation rises from 0.55 to 0.63; Indianapolis' 75<sup>th</sup> percentile rises from 0.49 to 0.59; Phoenix's median rises from 0.37 to 0.42; and Salt Lake City's 75<sup>th</sup> percentile rises from 0.33 to 0.51.



one in our data since we exclude within-MSA moves and no MSAs are perfectly correlated with each other.)

For example, in New York, the 75<sup>th</sup> percentile correlation shifts up to 0.95 from 0.53 in Figure 1, and the 25<sup>th</sup> percentile shifts to 0.50 from 0.12. In Miami and Detroit, the correlation is expected to be at least 0.66 three-quarters of the time. In most cities, even the 5<sup>th</sup> percentile correlation rises considerably. In New York, it increases to 0.31 from -0.22. Similarly large shifts of the entire probability distribution occur in San Francisco, Dallas, and Philadelphia, among others.

Not every metro area experiences such a sizeable shift in their correlation distributions after weighting. Atlanta and Phoenix, for example, change relatively little. Also, in many MSAs the 5<sup>th</sup> percentile correlation remains fairly low or negative. While those MSAs still exhibit an overall upwards shift in their distributions, the top three quartiles of the distributions shift by more than the bottom quartiles, as in Austin or Seattle.

To illustrate how high the effective expected correlations in real annual house price growth can be, Table 1 presents the distributions of the correlations across MSA-pairs. To convert the raw correlations to expected correlations, we weight each MSA-pair observation by the imputed probability of a household moving between those two MSAs. For comparison, the first column of Table 1 assumes that households have an equal probability of moving across any MSA-pair. This column corresponds to Figure 1 with all MSAs pooled together. In the equally-weighted case, the average pairwise correlation is just 0.34.<sup>8</sup> However, 25 percent of the MSA-pairs have correlations of at least 0.54. Column (2) uses population-weighted correlations, where the odds of a household moving to an MSA, conditional on moving at all, is proportional to that MSA's share of the total population in all MSAs. This assumption makes little difference

---

<sup>8</sup> We use equal weights for consistency with prior research. See, for example, Glaeser and Gyourko (2006).

relative to column (1): The median correlation rises from 0.35 to 0.39 and the 95<sup>th</sup> percentile correlation rises from 0.82 to 0.86.

The expected correlation in house price growth rises considerably across the distribution when we recognize that households tend to move between highly covarying MSAs. In column (3), we weight the MSA-pair correlations by the actual rate at which households moved between them, computed using the Treasury's county-to-county migration data described above. This column corresponds to Figure 2. Using such migration weights, the average correlation rises to 0.57 from 0.34 in column (1), and the median correlation rises to 0.60 from 0.35. An even more striking finding is that the 75<sup>th</sup> percentile correlation rises from 0.54 to 0.89 – i.e., 25 percent of household moves are between MSAs that have correlations in annual house price growth of 0.89 or above. Even the lower end of the distribution has remarkably high correlations. The 25<sup>th</sup> percentile correlation in column (3), 0.34, is about the same as the median correlation in column (1).

Overall, accounting for where households actually move when constructing expected correlations shifts the entire distribution of expected correlations to the right. This conclusion is confirmed by the bottom panel of Table 1 which reports various summary statistics for the (within-MSA) distribution of correlations. When MSA-pairs are weighted by migration flows, the difference between the 95<sup>th</sup> and 75<sup>th</sup> percentile correlations within each MSA shrinks dramatically, from 0.26 in column (1) to 0.08 in column (3), because the distribution is capped at 1.0 as it shifts to the right. Thus, the 75<sup>th</sup> percentile value rises considerably, but the 95<sup>th</sup> percentile cannot rise by as much. However, neither the interquartile range nor the difference between the 25<sup>th</sup> and 5<sup>th</sup> percentiles changes much between columns (1) and (3), indicating that the other points in the distribution increase in near-lockstep.

The next three columns of Table 1 show that the propensity to move between highly-covarying MSAs is not driven by relative house price changes in those markets. In the fourth column, we adjust the cross-MSA mobility rates by regressing the moving rate between a MSA-pair in a given year on the difference between the two MSAs in annual house price growth over the prior year, that difference squared, and a constant term. We then use the residuals from that regression, plus the constant term, as the moving weights.<sup>9</sup> The basic pattern of results is little changed from the one using the unadjusted moving rates, in the third column. The median expected covariance is 0.58 and the 95<sup>th</sup> percentile is 0.96. Although the 75<sup>th</sup> percentile covariance drops from 0.89 to 0.79, it is still considerably larger than for the unweighted or population-weighted moves in the first two columns.

In the fifth and sixth columns, we check whether the moving patterns are different in periods of house price growth or decline. They do not appear to be. The fifth column restricts the sample to MSAs in years where they had experienced positive nominal house price growth (“booms”). The sixth column restricts the sample to MSAs in years where they did not (“busts”). In both cases, the basic pattern of results is little changed from column (3).

Finally, we also compute correlations in five-year house price growth rates, rather than annual growth rates, to try to more closely match the average household’s holding period. As in column (3), we use raw household mobility patterns as the weights in forming expected correlations. The distribution of expected correlations, reported in the last column, rises across the board since using the longer-difference growth rates reduces the effect of high-frequency fluctuations or noise in house prices. 75 percent of households have house price correlations of

---

<sup>9</sup> This is akin to setting the price differences between the two MSAs to zero. Negative predicted moving rates are truncated at zero. While the estimated coefficients on the price difference terms are statistically significant and of an intuitive sign – households are less likely to move to an MSA whose price growth was higher than their own, and that effect accelerates as the price growth gap gets bigger – the magnitude is economically small and the R-squared is just 0.0013.

at least 0.44, 50 percent have correlations of at least 0.74, and 25 percent of households have correlations of 0.94 or more. Our ability to compute growth rates over even longer horizons is limited because of the length of the sample. However, the distribution using 10-year growth rates in house prices looks very similar to that reported in the last column of Table 1.

The reason that the expected correlation in house price growth rises when we account for households' moving tendencies is that the relation between household moves and house price correlation is skewed. The set of city pairs with high gross migration flows includes mostly city pairs with high price correlations, whereas the city pairs with low gross migration flows include both low and high correlation pairs. This pattern can be seen in Figure 3, which graphs the kernel density of the correlations in annual house price growth among various subsets of the 18,090 pairs of the 135 cities. The solid black line restricts the sample to the 4,812 city pairs that, according to the Treasury data, experienced no (gross) migration. A few of these pairs had correlations below  $-0.5$ , some had correlations of nearly 1, and the peak of the distribution is around 0.3. The dashed line shows the kernel density of the correlations for the remaining city pairs that experience positive gross migration. This distribution is slightly to the right of the distribution for city pairs with no migration, so on average households who move migrate between more highly correlated housing markets, but the overall difference is small. By contrast, the dotted line considers only those city pairs  $(A,B)$  where, of the households who moved out of city  $A$ , at least 1 percent moved to city  $B$ . One percent of moving households is a high moving rate, since most city pairs have few moves. The distribution of correlations shifts considerably to the right. That rightward shift is even more pronounced for the dash-dot line, which restricts the set of city pairs to those where the rate of moving out of  $A$  to  $B$  was at least

2.5 percent. These last two kernel densities indicate that the rate of moving is high mainly between high-correlation city pairs.

The patterns in Figure 3 explain the differences between Figures 1 and 2. When city pairs are equally weighted, the distribution of correlations looks like the solid and dashed lines. But when city pairs are weighted by the probability of a household moving between the two cities, the pairs represented by the solid line get no weight, while the pairs represented by the dotted and dash-dot distributions get significant weight. This reweighting shifts the mass of the correlation distribution (in most cities) to the right.

Investigating why MSAs have correlated house price growth rates – or why households tend to move between MSAs with correlated housing markets – is beyond the scope of this paper. Indeed, the value, to an individual household, of owning a house as a hedge against moving risk is separable from the reasons for the correlation or the potential move. For example, one possible reason for moving is that the shocks that induce the correlation in prices also induce labor market flows. That is, cities that are similar enough for households to want to move between are also likely to share the same economic fundamentals, leading to a correlation in their housing markets. Another possible reason is that cities with correlated house prices are demand substitutes. [McDuff 2008] In either case, for an individual household, owning a house still helps hedge the purchase price risk in the city that is moved to. Nonetheless, in our empirical section, below, we will confirm this conclusion by using exogenous predictors of cross-MSA mobility to form expected correlations and covariances.

### 3. *Data*

We use household level data from the 1980, 1990, and 2000 waves of the Integrated Public Use Microsample of the U.S. Census (IPUMS) as our base data set. The IPUMS is a representative cross-section sample of U.S. residents drawn from the decennial Census. It is well-suited for our purpose because it has many observations (in the 5 percent sample that we use, there are nearly 38 million person-level observations total in the three waves), contains MSA identifiers, reports whether households own their homes, and contains a host of household-level income and demographic characteristics that we use as controls. We construct two subsamples using this data, using the procedure outlined in Appendix Table C. Both subsamples are restricted to household heads age 25 and over, among other conditions. One sample, which we call the ‘migration’ subsample, excludes the households from the 1980 Census that due to the design of that year’s Census were not asked about prior residences. After these restrictions, 10.3 million household-level observations remain. The other sample, which we call the ‘regression’ sample, excludes households who either do not live in an MSA or live in MSAs that are not covered in our other data sets (as well as other, relatively minor, sample restrictions), for a net total of 4.2 million household-level observations.

The main variables of interest – the expected covariances ( $\text{cov}$ ), rent risks ( $s_r^2$ ), and expected lengths-of-stay ( $N$ ) – need to be imputed into the IPUMS. The expected covariance for a household living in MSA  $k$  is comprised of two parts: the vector of house price covariances between MSA  $k$  and the other MSAs, and the probability weights that a given household living in MSA  $k$  would apply to the likelihood of moving to each of the other MSAs:

$$E(\text{cov}(P_A, P_B))_{k,g} = \sum_l w_{k,l,g} \text{cov}(P_k, P_l). \text{ The house price covariance vector, } \text{cov}(P_k, P_l), \text{ is}$$

constructed by computing the covariance of real annual house price growth over the 1980 – 2002 period based on the OFHEO index described earlier.

We use several different approaches to estimate the moving weights,  $w_{k,l,g}$ . The IPUMS reports households' current MSA of residence and their MSA of residence five years earlier. Using the migration sample, we construct the average annual rate of moving from each MSA to each of the other MSAs conditional on moving out of the origin MSA, weighting the observations by the IPUMS household weights. We also repeat the exercise allowing the MSA-to-MSA moving matrix to differ by group,  $g$ . As groups, we use the Census's 381 detailed industries (1990 definition) as well as the Census's 243 detailed occupations (1990 definition). When we allow the weights to vary by MSA and group,  $E(\text{cov}(P_A, P_B))_{k,g}$  varies both across and within MSAs.

We construct expected horizon and rent volatility in the same manner as Sinai and Souleles (2005). We proxy for the expected horizon with the probability of not moving, imputed using exogenous demographic characteristics. The IPUMS reports whether a household has moved in the last year. To generate the expected probability of moving, we take the average rate of moving over the last year in the age (in 10-year bins)  $\times$  marital status  $\times$  occupation cell that matches the household in question (excluding the household from the cell). We subtract that average from one to obtain the probability of staying,  $P(\text{stay})$ . This is our inverse proxy for the expected horizon  $N$ . In the regression analysis, we will control separately for age, marital status, and occupation in the vector of demographic controls, so the probability of staying will be identified from households having a different mobility profile over their lifetimes depending on their marital status and occupation.

To estimate rent volatility, we use data from REIS, a commercial real estate data provider that has surveyed 'Class A' apartment buildings in 44 major markets between 1980 and the present. We use their measure of average effective rents by MSA, deflated using the CPI less

shelter. To estimate the volatility, we de-trend the log annual average real rent in each MSA and compute the standard deviation of the deviations from the trend between 1980 and 2002. By using logs, the standard deviation is calculated as a percent of the rent and so the measured rent risk is not affected by the level or average growth rate of rents.

#### 4. Estimation strategy and results

We wish to estimate the following regression, for household  $i$  in MSA  $k$  at time  $t$ :

$$\begin{aligned} OWN_{i,k,t} = & \beta_0 + \beta_1 f(s_r)_k + \beta_2 g(N)_i + \beta_3 f(s_r)_k \times g(N)_i + \\ & \beta_4 E(\text{cov}(P_A, P_B))^{\frac{1}{2}}_{i,k} + \beta_5 E(\text{cov}(P_A, P_B))^{\frac{1}{2}}_{i,k} \times g(N)_i + \\ & \theta X_i + \psi Z_{k,t} + \zeta_t + \varepsilon_{i,k,t}, \end{aligned} \quad (5)$$

where ‘OWN’ is an indicator variable for home ownership,  $s_r$  is the measure of the local rent volatility,  $N$  is the imputed probability of not moving (P(stay)), and  $E(\text{cov}(P_A, P_B))^{\frac{1}{2}}$  is the square root of the household’s expected covariance. The remaining variables are controls:  $X$  is a vector of household-level characteristics,  $Z$  controls for time-varying MSA-level characteristics, and  $\zeta$  is a vector of year dummies. We will estimate equation (5) using OLS and linear IV.

The main predictions from the framework in Section 1 are that  $\beta_3 > 0$ ,  $\beta_4 > 0$ , and  $\beta_5 < 0$ . The tests of  $\beta_4 > 0$  and  $\beta_5 < 0$  are novel to this paper, while  $\beta_3 > 0$  was tested in Sinai and Souleles (2005) using a different data set. The framework in Section 1 also shows that the probability of owning should increase with the horizon  $N$ , which implies that  $\beta_2 > 0$ . In practice the coefficient on the uninteracted horizon term can also pick up the effects of other factors affecting the probability of owning, such as fixed costs being amortized over longer stays, and so we will not emphasize  $\beta_2$ . Also, we will include MSA  $\times$  year dummies, which will subsume the effect of  $\beta_1$ , the coefficient on the uninteracted rent volatility.



Most of our empirical focus will be on identifying the estimates of  $\beta_4$  and  $\beta_5$ . One way to identify  $\beta_4$ , the effect of the expected covariance on the decision to own, is to make use of the fact that different types of households within each MSA can have different weights applied to the MSA  $\times$  MSA house price covariance matrix, reflecting their different probabilities of moving to different MSAs. This yields variation in *expected* covariances within *and* across MSA. Since we allow the moving weights to differ by industry or occupation groups, the resulting variation is at the industry  $\times$  MSA or the occupation  $\times$  MSA level. Thus we can include MSA  $\times$  year dummies and a complete set of industry or occupation dummies, and still identify  $\beta_4$ . If we assume instead that all households within an MSA have the same probabilities of moving to the other MSAs,  $\beta_4$  cannot be separately identified from the MSA  $\times$  year fixed effects.

To illustrate, we could compare the homeownership rates of electricians and lawyers in Philadelphia since those two industry groups, despite living in the same MSA, are likely to move to different cities and thus have different expected covariances. We can then compare the difference in the two groups' homeownership rates to the difference in homeownership rates in the same two industry groups in another city, such as Seattle, who themselves have different expected covariances. The fact that households may be on average more likely to own their houses in Seattle or Philadelphia is absorbed with MSA  $\times$  year dummies. Any capitalization of the value of the moving hedge into house prices is also captured by these fixed effects. Average differences in homeownership rates across industries can be controlled for with industry dummies. Thus our estimate of  $\beta_4$  would be identified by industry  $\times$  MSA variation. Our identification strategy generalizes this example to all MSAs and all industries. We also try a similar strategy using occupation groups rather than industries. We allow for general non-

independence of the standard errors within industry  $\times$  MSA  $\times$  year (or occupation  $\times$  MSA  $\times$  year) groups by clustering the standard errors.

We can identify  $\beta_5$  separately from  $\beta_4$ , since  $\beta_5$  is the coefficient on the interaction between the probability of staying ( $N$ ) and the  $E(\text{cov}(P_A, P_B))^{1/2}$  term. Whether or not the expected covariance term alone is distinguishable from unobserved MSA heterogeneity, the probability of staying varies across households  $i$  based on their demographic characteristics, so  $E(\text{cov}(P_A, P_B))^{1/2}_{i,k} \times g(N)_i$  varies by household within MSA. The same logic applies to the identification of  $\beta_3$ , the coefficient on  $f(s_r)_k \times g(N)_i$ , since the interaction of the MSA level rent variance and the group-level probability of staying provides MSA  $\times$  group variation.

In addition to the variables of interest, in all regressions we control for standard household-level covariates: The (uninteracted) imputed P(stay), log real family income, age dummies (in 10-year bins), indicator variables for marital status (single, married, widowed, and divorced), education dummies (less than high school, high school, some college, and completed college), 381 occupation dummies, and the probability of moving within the MSA conditional on moving. We add 243 industry dummies when we use industry for the group  $g$  in forming the moving weights. MSA-level covariates and aggregate time series effects are subsumed by the MSA  $\times$  year fixed effects, which we include in all specifications.

As a baseline, the first column of Table 2 uses the average MSA-to-MSA mobility rates by occupation to construct the moving probabilities, so  $E(\text{cov}(P_A, P_B))^{1/2}$  varies both across and within MSAs. The predictions of the framework from Section 1 are supported by the data. In column (1), the first row reports the effect of the expected price covariance on the likelihood of owning and the second row reports how the effect differs as the probability of staying increases.

Because the specification includes an interaction term with P(stay), the coefficient in the first row can be interpreted as the effect of the covariance for the polar case household who expects to move right away. We find that such a household is more likely to own when the covariance is greater, as predicted. The estimated coefficient  $\beta_4$  of 2.042 (0.252 standard error) implies that a one standard deviation increase in the square root of the expected covariance (0.015 on a base of 0.041) would yield a 3.1 percent increase in the probability of home ownership ( $2.04 \times 0.015$ ). This represents a sizeable increase in the ownership rate (the average is 65.4 percent), though a small fraction of the cross-sectional standard deviation in the likelihood of owning (the standard deviation is 47.6 percent). However, this extrapolation to an expected horizon of zero years is well outside of the variation in the sample.

To assess the effect for households with a longer expected horizon, we need to turn to the interaction term,  $P(\text{stay}) \times E(\text{cov}(P_A, P_B))^{\frac{1}{2}}$ , in the second row. The negative coefficient  $\beta_5$  of  $-2.039$  (0.283) implies that the effect of the price covariance on the probability of owning attenuates as a household's expected horizon increases and the value of the moving-hedge falls, again as predicted. Since the estimated coefficient on the interaction term is of similar magnitude but opposite sign to the coefficient on  $E(\text{cov}(P_A, P_B))^{\frac{1}{2}}$ , the net effect of higher covariance declines to effectively zero for the polar case household who is imputed to never move, i.e. whose  $P(\text{stay})=1$ . This result, too, is consistent with the framework in Section 1, since non-movers have no need for a moving hedge. To be precise, the estimates show that households with short expected durations and high expected covariances are more likely to own their houses than otherwise equivalent households with short expected durations and low expected covariances in the same MSA and year. As the expected duration increases, the gap in the probability of ownership declines to zero.

We can combine the estimates for these two polar cases to estimate the effect of a one standard deviation higher expected covariance on the probability of owning for households with an intermediate expected duration. For a household with the lowest expected horizon in the sample, two years (imputed  $P(\text{stay})=0.5$ ), a one standard deviation higher expected covariance raises the probability of owning by 1.6 percentage points (0.2 percentage points standard error), which is a large and significant effect. Households with a longer, five-year expected horizon (with  $P(\text{stay})=0.8$ , approximately the sample average), are 0.63 percentage points (0.15) more likely to own a house when the expected covariance rises by one standard deviation. By the time households have a one-in-ten chance of moving in any given year ( $P(\text{stay})=0.9$ ), the effect on the probability of owning approaches zero (0.32 percentage points) and is barely statistically significant (a standard error of 0.16).

In the third row of Table 2, the effect of rent volatility in the current MSA on the likelihood of owning becomes more positive as the horizon increases, since the coefficient on the interaction term,  $P(\text{stay}) \times s_r$ , is positive. This interaction effect was one of the main empirical results in Sinai and Souleles (2005), and it is robust throughout all the specifications in this paper.

We include  $P(\text{stay})$  separately to control for expected horizon and related factors like the fixed costs of buying or selling a house. The estimated coefficient is positive as expected: households with longer expected lengths of stay on average are more likely to own.

The second column of Table 2 reports the estimated coefficients when we impute the pattern of where households expect to move based on industry groups and MSA. The results are very similar to those using occupation groups.

In the third column, we eliminate the within-MSA group-based variation by imputing mobility based on where other households in the same MSA move. Because the weighting matrix in this specification varies only by MSA, the estimated coefficients on the uninteracted expected covariance (and rent volatility) term cannot be identified separately from the unobserved MSA-level heterogeneity that is absorbed by MSA  $\times$  year effects. In this case, only the interaction terms with expected horizon (P(stay)) can be identified, since they vary by MSA  $\times$  household. Despite the difference in the source of variation, the estimated coefficients on the interaction terms are very close to those in columns (1) and (2).

The last two columns of Table 2, by leaving out the interaction term P(stay)  $\times$   $E(\text{cov}(P_A, P_B))^{\frac{1}{2}}$ , provides the average effect of the expected covariance on home owning, across households of all expected durations. These estimates are identified solely from the within-MSA variation (by occupation or industry) and are not dependent on the P(stay) imputation. Because  $E(\text{cov}(P_A, P_B))^{\frac{1}{2}}$  in this specification is not interacted with P(stay), the estimated elasticity corresponds to the household with the average expected length-of-stay in the sample, which is about five years (P(stay)=0.80). The estimated coefficient in the top row of column (4) is 0.367 (0.99), which suggests that if the expected covariance were to rise by one standard deviation, the probability of home owning would go up by  $0.367 \times 0.015$ , or almost 0.6 percentage points. Similar results are found using industry groups in column (5).

One potential complication is that if households are constrained to move primarily to where they can afford to move, they might have to move to markets that have correlated house price changes. If this effect happens to be stronger when the homeownership rate is higher, it could potentially spuriously link the probability of owning with the expected covariance, since our measure of expected covariance is constructed using actual moves. However, there is little

reason to think that this effect would be stronger within MSAs for the industries or occupations that have a higher homeownership rate, and so it should not affect our analysis.<sup>10</sup>

To confirm that this mechanism isn't driving the results, we instrument for our measure of expected covariance with a covariance constructed using mobility weights intended to proxy for where a household would like to move if it were unconstrained. For this robustness check, we assume that, conditional on moving, households would like to move to those cities where other households in their current occupation or industry tend to locate. In particular, we impute household  $i$ 's (currently living in MSA  $k$ ) probability of moving to MSA  $l$  as simply the share of household  $i$ 's occupation (or industry) in MSA  $l$  relative to all other MSAs excluding  $k$ . (E.g., the instrument is constructed such that, if a large fraction of the nation's lawyers live in Philadelphia, then lawyers in other cities have a proportionally higher probability of moving to Philadelphia.) This set of probabilities differs by occupation (industry) within an MSA, but most of the variation in this instrument is across occupations (industries), not across MSAs. However, when those moving probabilities are used to weight the MSA covariance vectors, which differ across MSAs, the resulting expected covariance instrument varies both within and across MSAs. (The first stages of the IV regressions are reported in Appendix Table B.)

The results of this IV strategy are reported in Table 3. The specification in column (1) corresponds to the first column of Table 2, except now we instrument for the expected covariance using the weighted covariance described above that assumes that households move according to the distribution of their occupation shares, which varies by occupation  $\times$  MSA. To instrument for  $P(\text{stay}) \times E(\text{cov}(P_A, P_B))^{\frac{1}{2}}$ , we interact  $P(\text{stay})$  with the expected covariance

---

<sup>10</sup> The theoretical argument of this paper does not require that a household's moving between markets be unconstrained. If a household knows in advance that, when it moves, it would need to move to a housing market where house prices had covaried positively, it should take that into account when making its initial house purchase decision. This is why we address just the empirical implications of this potential mechanism.

instrument. Since  $P(\text{stay})$  is imputed using exogenous household demographics, we do not need an instrument for it.

The IV-estimated coefficients on  $E(\text{cov}(P_A, P_B))^{\frac{1}{2}}$  and its interaction with  $P(\text{stay})$  nearly double relative to the OLS estimates in Table 2, to 4.209 (1.271) for the expected covariance and  $-3.813$  (0.387) for the interaction term. Despite the larger standard errors, the IV coefficients remain statistically significant. Since the estimates continue to be close in magnitude with opposite signs, at every expected length-of-stay the estimated effect of higher expected covariance on the probability of homeownership is about twice what was found in Table 2. As expected, those households who we impute would never move still have no differential response to changes in expected covariance. For households with  $P(\text{stay})=0.5$ , a one standard deviation greater expected covariance leads to a nearly 4 percentage point increase in the probability of home owning. At the sample average  $P(\text{stay})$ , 0.8, the same increase in expected covariance would raise the probability of owning by just under 2 percentage points. Despite the statistical significance of the individual reported coefficients, for levels of  $P(\text{stay})$  observed in the data, the combined effect of  $E(\text{cov}(P_A, P_B))^{\frac{1}{2}}$  and  $E(\text{cov}(P_A, P_B))^{\frac{1}{2}} \times P(\text{stay})$  becomes statistically indistinguishable from zero when instrumenting.

Column (2) of Table 3 uses industry groups rather than occupation to construct the instrument. The estimated coefficients on  $E(\text{cov}(P_A, P_B))^{\frac{1}{2}}$  and  $E(\text{cov}(P_A, P_B))^{\frac{1}{2}} \times P(\text{stay})$  are slightly smaller in absolute value than in column (1), yet still statistically significant. Their combined effects are also statistically different from zero as long as  $P(\text{stay})$  is lower than 0.65. At  $P(\text{stay})=0.5$ , a one standard deviation greater expected covariance leads to a 3 percentage point increase in the probability of owning.

Column (3) uses just the MSA-level variation to construct the expected covariance, and for the instrument uses the occupation distribution of the MSA to construct where the MSA residents on average would move to. As in Table 1, the level effect of expected covariance cannot be identified in this particular specification. However, the interaction term has the same estimated magnitude as in columns (1) and (2), with about a 50 percent smaller standard error. In sum, the interaction term remains consistently significantly and economically negative, even across all the IV specifications. Thus we can reject the null hypothesis of no attenuation (with expected horizon) of the effect of the moving hedge on the probability of owning on average over the sample.

## 5. *Conclusion*

This paper established two novel results. First, because households tend to move among correlated housing markets, the effective covariance of house prices across housing markets is much higher than analysts have previously assumed. We find that half of households' moves in the U.S., excluding within-MSA moving, are between MSAs with correlations in annual real house price growth rates of 0.60 or greater, and 25 percent are between MSAs with more than a 0.89 correlation. When five-year house price growth rates are used, the median correlation rises to 0.74.

Second, households' tenure decisions appear to be sensitive to the moving-hedge benefit of home owning. Households with higher expected covariances between house prices in their current market and their possible future markets are more likely to own. This effect attenuates with a household's expected length of stay in the house: less mobile households place less weight on future housing markets than do more mobile households. For a household who is



likely to move in the next two years, having a one standard deviation higher expected covariance leads to a 2 to 4 percentage point higher homeownership rate. That relationship diminishes with the expected length of stay in the house and is indistinguishable from zero for households who appear unlikely to move. The results are robust to instrumenting for the expected covariances by assuming households move proportionally to their industry shares.

The analysis in this paper suggests that the natural hedge provided by owning a house may be quite valuable, and for many households home owning (absent leverage) may actually reduce the risk of the total lifetime cost of obtaining housing services. As noted in Sinai and Souleles (2005), home owners who expect never to move are in effect hedged against housing cost risk, since they locked in their housing costs with their initial house purchase. In this paper, we find that even home owners who expect to sell their house and move to another market may be partially hedged, because the volatility in their current house price often undoes the volatility in the price of their next house, reducing their lifetime risk on net. By contrast, renters are exposed to volatility in housing costs in both their current market and any future markets they might move to.

The argument that households should manage the uncertainty of their total lifetime housing costs, not the volatility of their current house price, has important implications for evaluating the efficacy of various methods of controlling households' housing risk, including house price derivatives. For most households, the positive expected covariance between house prices in their current city and prices in possible future cities provides at least a partial hedge against house price risk when they move. Because of that, households who use housing derivatives to lock in the sale price of their current house may actually unhedge themselves by reducing the covariance to zero. Instead, households would need to obtain a more complex

portfolio of derivatives that would hedge their total housing costs, including their future housing costs in different markets. This important distinction is neglected in analyses that implicitly assume that the covariances between the current and potential future housing markets are low (e.g., Case *et al.* (1993), Geltner *et al.* (1995), Shiller (2008), Voicu (2007)).<sup>11</sup>

This analysis may help explain why the house price futures market has failed to take off. [Shiller (2008)] It may simply be less expensive, easier, and nearly as effective to hedge by owning a home. While the natural hedge provided by owning a house is, in most cases, a partial one, providing a supplemental financial hedge against total housing risk might be complex. This analysis may also help explain why there are so few long-term leases in the U.S. [Genesove (1999)] A long-term lease avoids rent risk and leaves the sale price risk with the landlord. But for many mobile households there is a potential benefit of retaining exposure to the sale price, as a hedge against the uncertainty of the cost of future housing services.

Lastly, the results in this paper suggest that, for many households, the marginal propensity to consume out of changes in house prices could be small. To illustrate, if an increase in house prices reflects the fact that a household's implicit short position in future housing costs has become commensurately more expensive, then the household's real wealth is effectively unchanged by the housing capital gains. Unless the capital gains alleviate liquidity/collateral constraints, the consumption response should accordingly be small. Prior research emphasized this argument for home owners with long expected stays in their current housing markets. [Sinai and Souleles (2005), Campbell and Cocco (2007)] In this paper, the high expected house price covariances that we find imply that even home owners with short expected stays in their current housing markets can face similarly small real wealth effects as home owners with long horizons.

---

<sup>11</sup> de Jong *et al.* (2007) point out another reason that housing derivatives might provide a poor hedge is that MSA-level house price indices do not explain much of the variation in individual house prices. This point applies to within-MSA moves as well.

That is, the high covariances effectively lengthen the horizon of households that are likely movers, and so the wealth effect from a change in house prices will often be largely offset by changes in housing costs in both current and expected future housing markets. These results can help explain the small marginal propensities to consume out of housing wealth found by Calomiris *et al* (2009), Attanasio *et al* (2009), and by Campbell and Cocco (2007) for young homeowners, and suggest that the main channels for housing wealth effects on non-housing consumption would have to be through affecting collateral or liquidity constraints. [Lustig and Van Nieuweburgh (2005), Ortalo-Magné and Rady (2006)]

## References

- Attanasio, Orazio, Laura Blow, Robert Hamilton, and Andrew Leicester. 2009. 'Booms and Busts: Consumption, House Prices and Expectations.' *Economica*, 76, 20-50.
- Calomiris, Charles, Stanley Longhofer, and William Miles. 2009. 'The (Mythical?) Housing Wealth Effect.' *NBER Working Paper #15075*, June.
- Caplin, Andrew, William Goetzmann, Eric Hangen, Barry Nalebuff, Elisabeth Prentice, John Rodkin, Matthew Spiegel and Tom Skinner. 2003. 'Home Equity Insurance: A Pilot Project,' Working Paper No. ysm372, Yale School of Management, Yale University.
- Case, Karl, and Robert Shiller. 1989. 'The Efficiency of the Market for Single-Family Homes.' *American Economic Review*, 79 (1), 125-137.
- Case, Karl, Robert Shiller, and Allan Weiss. 1993. 'Index-Based Futures and Options Markets in Real Estate,' *Journal of Portfolio Management*, 19(2), 83-92.
- Chan, Sewin. 2001. 'Spatial Lock-In: Do Falling House Prices Constrain Residential Mobility?' *Journal of Urban Economics*, 49(3), 567-586.
- Cocco, Joao. 2000. 'Hedging House Price Risk with Incomplete Markets,' paper presented at the American Finance Association 2001 Meetings, New Orleans.
- Davidoff, Thomas. 2006. 'Labor Income, Housing Prices and Homeownership,' *Journal of Urban Economics*, 59(2), 209-235.
- De Jong, Frank, Joost Driessen, and Otto Van Hemert. 2007. 'Hedging House Price Risk: Portfolio Choice with Housing Futures,' *mimeo*, New York University.
- Geltner, David, Norman Miller, and Jean Snavely. 1995. 'We Need a Fourth Asset Class: HEITs,' *Real Estate Finance*, 12(2), 71-81.
- Genesove, David. 1999. 'The Nominal Rigidity of Apartment Rents,' Working Paper No. 7137, National Bureau of Economic Research.
- Genesove, David and Christopher Mayer. 1997. 'Equity and Time to Sale in the Real Estate Market,' *The American Economic Review*, 87(3), 255-69.
- Genesove, David and Christopher Mayer. 2001. 'Loss Aversion and Seller Behavior: Evidence from the Housing Market,' *Quarterly Journal of Economics*, 116(4), 1233-1260.
- Glaeser, Edward and Joseph Gyourko. 2006. 'Housing Dynamics,' Working Paper No. 12787, National Bureau of Economic Research.

- Han, Lu. 2008. 'The Effects of House Price Uncertainty on Housing Demand: Empirical Evidence from the U.S. Housing Markets,' *Review of Financial Studies*, forthcoming.
- Hurst, Erik and Frank Stafford. 2004. 'Home Is Where the Equity Is: Mortgage Refinancing and Household Consumption,' *Journal of Money, Credit & Banking*, 36(6), 985-1014.
- Li, Wenli and Rui Yao. 2007. 'The Life-Cycle Effects of House Price Changes,' *Journal of Money, Credit and Banking*, 39(6), 1375-1409.
- Lustig, Hanno and Stijn Van Nieuwerburgh. 2005. 'Housing Collateral, Consumption Insurance, and Risk Premia: An Empirical Perspective,' *Journal of Finance*, 60(3), 1167-1219.
- Meese, Richard and Nancy Wallace. 1994. 'Testing the Present Value Relation for Housing Prices: Should I Leave my House in San Francisco?,' *Journal of Urban Economics*, 35, 245-266.
- Ortalo-Magné, Francois and Sven Rady. 2002. 'Tenure Choice and the Riskiness of Non-Housing Consumption,' *Journal of Housing Economics*, 11, 266-279.
- . 2006. 'Housing Market Dynamics: On the contribution of income shocks and credit constraints.' *Review of Economic Studies*, 73, 459-485.
- Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. 2008. *Integrated Public Use Microdata Series: Version 4.0*. Minneapolis: Minnesota Population Center.
- Shiller, Robert. 2008. 'Derivatives Markets for Home Prices,' Working Paper No. 13962, National Bureau of Economic Research.
- Sinai, Todd and Nicholas Souleles. 2005. 'Owner Occupied Housing as a Hedge Against Rent Risk,' *Quarterly Journal of Economics*, 120(2), 763-789.
- Stein, Jeremy. 1995. 'Prices and Trading Volume in the Housing Market: A Model with Downpayment Effects,' *Quarterly Journal of Economics*, 110, 379-406.
- Voicu, Cristian. 2007. 'Optimal Portfolios with Housing Derivatives,' *mimeo*, Harvard Business School.

Table 1: Distributions of correlations in house price growth across MSAs

Correlation:	One-year growth rates of real house prices						Five-year growth rates
	Migration weighted						
	Unweighted	Population weighted	Average migration rates	Conditional on house price growth differences	During booms	During busts	Average migration rates
Weighting scheme:							
Average	0.34	0.37	0.57	0.55	0.59	0.56	0.64
5 <sup>th</sup> percentile	-0.17	-0.16	-0.02	-0.004	0.04	-0.05	-0.06
25 <sup>th</sup> percentile	0.15	0.18	0.34	0.36	0.44	0.32	0.44
50 <sup>th</sup> percentile	0.35	0.39	0.60	0.58	0.64	0.60	0.74
75 <sup>th</sup> percentile	0.54	0.58	0.89	0.79	0.80	0.84	0.94
95 <sup>th</sup> percentile	0.82	0.86	0.97	0.96	0.96	0.97	0.99
Average difference: 95 <sup>th</sup> –75 <sup>th</sup> percentile	0.26	0.23	0.08	0.14	0.12	0.15	0.07
Average interquartile range	0.37	0.35	0.37	0.20	0.16	0.15	0.41
Average difference: 25 <sup>th</sup> –5 <sup>th</sup> percentile	0.27	0.27	0.33	0.39	0.31	0.33	0.45

Notes: This table reports the distribution of weighted average house price growth correlations across MSAs. MSA  $k$ 's weighted average correlation is between its real house price growth and that in all other MSAs, with weights varied as listed in the table. Correlations are constant over time within MSA pairs, but the weights can vary over time. A 'boom' is an MSA  $x$  year when nominal house price growth in MSA  $k$  was positive; a 'bust' is an MSA  $x$  year when nominal house price growth was negative or zero.

Table 2: The relationship between expected house price covariance and the probability of owning (OLS)

	(1)	(2)	(3)	(4)	(5)
$E[\text{cov}(P_A, P_B)]^{1/2}$	2.042 (0.252)	1.950 (0.265)		0.367 (0.099)	0.298 (0.115)
$P(\text{stay})_i \times E[\text{cov}(P_A, P_B)]^{1/2}$	-2.039 (0.283)	-2.016 (0.292)	-2.135 (0.151)		
$P(\text{stay})_i \times s_r$	2.430 (0.140)	2.468 (0.152)	2.464 (0.081)	1.856 (0.130)	1.902 (0.143)
$P(\text{stay})_i$	0.887 (0.019)	0.870 (0.019)	0.889 (0.010)	0.838 (0.017)	0.821 (0.016)
Probability weights for $E[\text{cov}(P_A, P_B)]$ :	MSA x occupation	MSA x Industry	MSA	MSA x occupation	MSA x Industry
MSA $\times$ year dummies?	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared	0.2965	0.2993	0.2964	0.2965	0.2993
Clustering of standard errors	MSA x year x occupation	MSA x year x industry	None	MSA x year x occupation	MSA x year x industry

Notes:  $N = 3,326,113$ . Sample period covers the 1980, 1990, and 2000 Censuses. The dependent variable is an indicator variable that takes the value of one if the respondent owns its home and zero if the respondent rents. The probability of staying (not moving),  $P(\text{stay})$ , is imputed using occupation  $\times$  marital status  $\times$  age category cells. The square root of the expected covariance of house prices,  $E[\text{cov}(P_A, P_B)]^{1/2}$ , is a moving-probability weighted average of the covariances between the MSA of residence and possible future MSAs. The standard deviation of detrended log rent,  $s_r$ , is a MSA ( $k$ ) characteristic and is subsumed by the MSA  $\times$  year dummies. The standard deviations and covariances are not time-varying. All regressions include as covariates: MSA  $\times$  year dummies, age dummies, occupation dummies, marital status dummies, education dummies, log real family income, and the share of moving households who remain in the MSA. Columns (2) and (5) add a detailed set of industry dummies.

Table 3: The relationship between expected house price covariance and the probability of owning (IV)

	(1)	(2)	(3)
$E[\text{cov}(P_A, P_B)]^{1/2}$	4.209 (1.271)	3.757 (0.743)	
$P(\text{stay})_i \times E[\text{cov}(P_A, P_B)]^{1/2}$	-3.813 (0.387)	-3.741 (0.377)	-3.718 (0.162)
$P(\text{stay})_i \times s_r$	2.911 (0.155)	2.946 (0.171)	2.908 (0.083)
$P(\text{stay})_i$	0.930 (0.021)	0.912 (0.021)	0.927 (0.010)
Probability weights for $E[\text{cov}(P_A, P_B)]$ :	MSA x Occupation	MSA x Industry	MSA
MSA $\times$ year dummies?	Yes	Yes	Yes
Instrument?	MSA x Occupation instrument	MSA x Industry instrument	MSA x Occupation instrument
Adjusted R-squared	0.2965	0.2993	0.2964
Clustering of standard errors	MSA x year x Occupation	MSA x year x industry	none

Notes:  $N = 3,326,113$ . Sample period covers the 1980, 1990, and 2000 Censuses. The dependent variable is an indicator variable that takes the value of one if the respondent owns its home and zero if the respondent rents. The probability of staying,  $P(\text{stay})$ , is imputed using occupation  $\times$  marital status  $\times$  age category cells. The square root of the expected covariance of house prices,  $E[\text{cov}(P_A, P_B)]^{1/2}$ , is a moving-probability weighted average of the covariances between the MSA of residence and possible future MSAs. The standard deviation of detrended log rent,  $s_r$ , is a MSA ( $k$ ) characteristic and is subsumed by the MSA  $\times$  year dummies. The standard deviations and covariances are not time-varying. The instrument replaces the actual moving rates between MSAs with each destination MSA's share of the occupation (industry). All regressions include as covariates: MSA  $\times$  year dummies, age dummies, occupation dummies, marital status dummies, education dummies, log real family income, and the share of moving households who remain in the MSA. Column (2) adds a detailed set of industry dummies.



Figure 1:

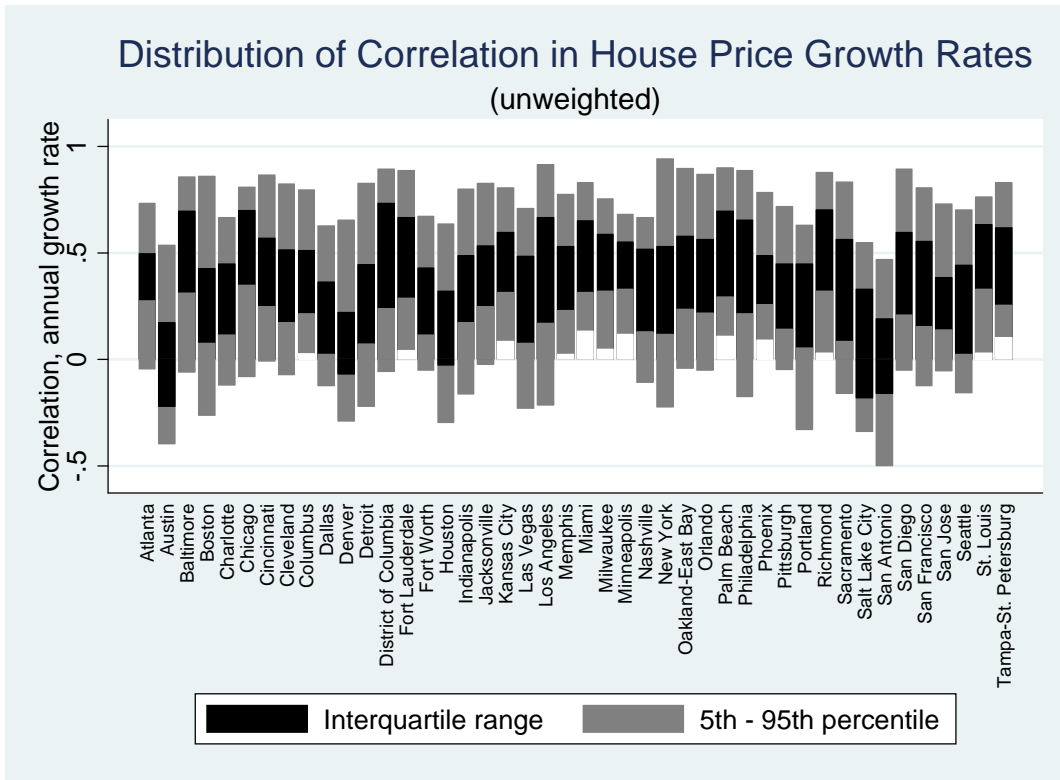


Figure 2:

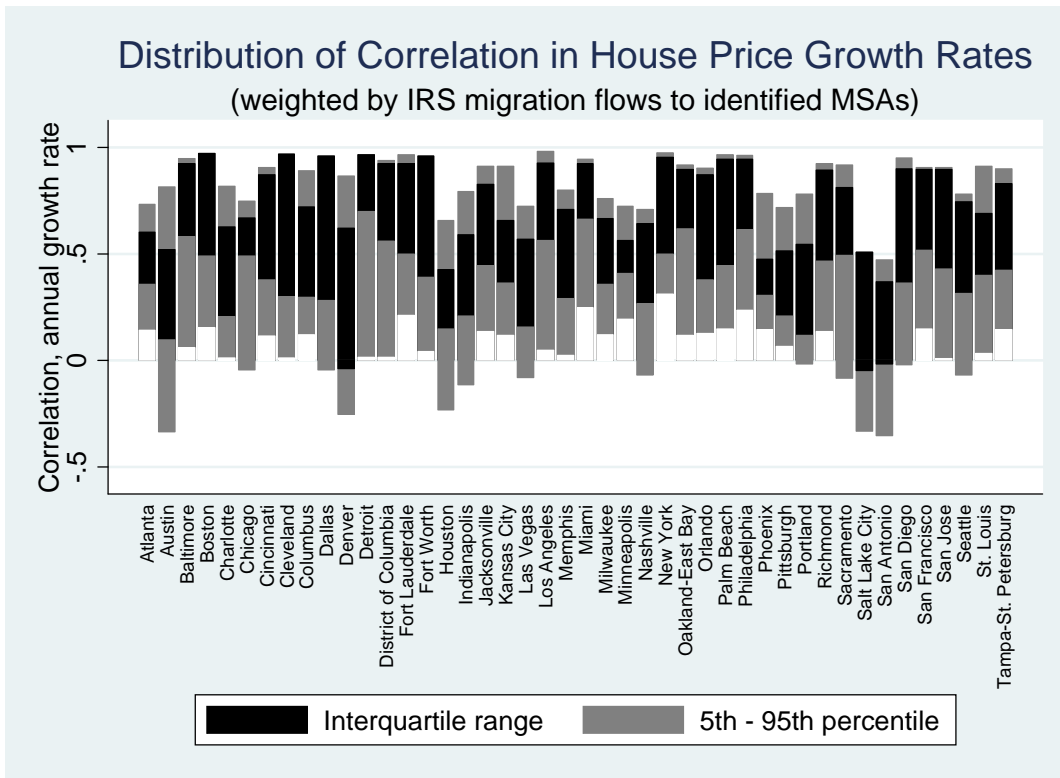
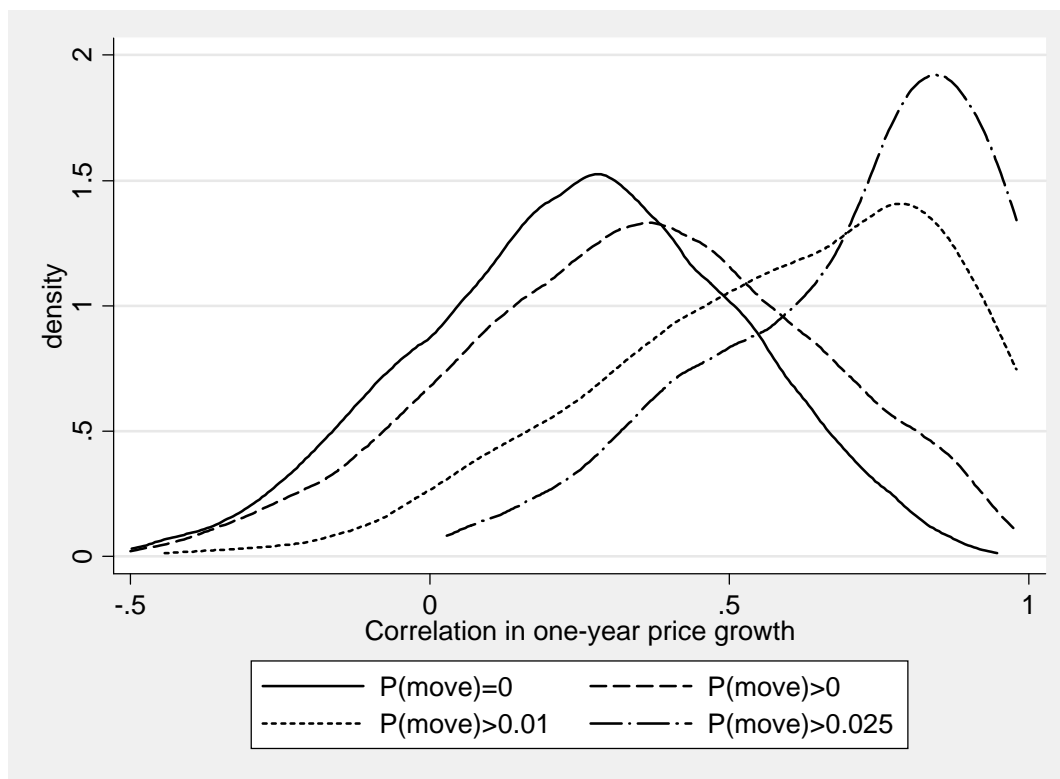


Figure 3: The distributions of house price growth correlations across housing markets, by probability of moving



Appendix Table A: Summary Statistics

Variable	Mean	Standard Deviation
Own	0.654	0.476
Probability of staying	0.810	0.113
SD(real rent growth)	0.068	0.028
SQRT(IPUMS MSA $\times$ industry-weighted average price covariance) – actual	0.041	0.015
SQRT(IPUMS MSA $\times$ industry-weighted average price covariance) – imputed	0.038	0.011
Average annual rent (\$2000)	10,290	3,738
Average house price (\$2000)	170,235	87,066
Rent growth rate (real)	0.011	0.008
Price growth rate (real)	0.017	0.013
Age	44.0	11.1
Family income (\$2000)	64,534	58,857
Fraction married	0.64	0.48
Fraction widowed	0.05	0.21
Fraction divorced	0.19	0.39
Fraction with less than high school ed.	0.16	0.37
Fraction with high school diploma	0.28	0.45
Fraction who attended some college	0.26	0.44
Fraction with college diploma	0.30	0.46
Share of moves that are within-MSA	0.75	0.04

Appendix Table B: First stages of the IV regressions (Table 3)

Column of Table 3:	(1)		(2)		(3)
Original covariate	$\frac{P(\text{stay})_i \times E[\text{cov}(P_A, P_B)]^{1/2}}{E[\text{cov}(P_A, P_B)]^{1/2}}$	$E[\text{cov}(P_A, P_B)]^{1/2}$	$\frac{P(\text{stay})_i \times E[\text{cov}(P_A, P_B)]^{1/2}}{E[\text{cov}(P_A, P_B)]^{1/2}}$	$E[\text{cov}(P_A, P_B)]^{1/2}$	$\frac{P(\text{stay})_i \times E[\text{cov}(P_A, P_B)]^{1/2}}{E[\text{cov}(P_A, P_B)]^{1/2}}$
Constructed $E[\text{cov}(P_A, P_B)]^{1/2}$	-0.950 (0.0003)	0.324 (0.002)	-0.681 (0.002)	0.393 (0.002)	-0.614 (0.001)
$P(\text{stay})_i \times$ constructed $E[\text{cov}(P_A, P_B)]^{1/2}$	1.173 (0.0002)	0.067 (0.002)	1.217 (0.001)	0.027 (0.001)	1.166 (0.001)
Adjusted R-squared	0.9984	0.9520	0.9558	0.9628	0.9651
Probability weights for $E[\text{cov}(P_A, P_B)]$ :	MSA x Occupation		MSA x Industry		MSA x Occupation
Instrument?	MSA x Occupation		MSA x Industry		MSA x Occupation
Clustering of standard errors	MSA x year x occupation		MSA x year x industry		none

Notes: N = 3,326,113. Sample period covers the 1980, 1990, and 2000 Censuses. The covariance of house prices is a probability-weighted average of the covariances between the MSA of residence and possible future MSAs. The dependent variable is the expected covariance weighted by the actual probability of staying (not moving). The instrument replaces the actual moving rates with each destination MSA's share of the industry or occupation group. The probability of staying is imputed using occupation  $\times$  marital status  $\times$  age category cells. The standard deviations and covariance are not time-varying. All regressions include MSA  $\times$  year dummies,  $P(\text{stay})$ , the standard deviation of real rents, age dummies, occupation dummies, marital status dummies, education dummies, log real family income, the share of moving households who remain in the MSA, and a detailed set of industry dummies.

Appendix Table C: IPUMS Data Construction

	Observations deleted	Observations left
Starting observations		37,925,632
Exclude age<25	14,159,970	23,765,662
Keep only heads of household	10,670,270	13,095,392
Keep only single-family households	877,100	12,218,292
Drop if were military households 5 years prior	61,092	12,157,200
Drop if were in college 5 years prior	192,144	11,965,056
<b><i>Migration sample:</i></b>		
Starting observations		11,965,056
Not asked about prior residence in 1980 Census	1,626,943	10,338,113
<b><i>Regression sample:</i></b>		
Starting observations		11,965,056
Cannot match MSA to REIS or HPI data <sup>1</sup>	6,765,722	5,199,334
Exclude MSAs with incomplete HPI data <sup>2</sup>	41,261	5,158,073
Missing log(family income)	64,204	5,093,869
Unable to impute P(stay)	494	5,093,375
Missing sd(rent)	764,579	4,328,796
Cannot impute out-of-town moves	22,830	4,281,924
Cannot impute expected moves based on occupation or industry shares	32,019	4,249,905
Drop if P(stay)<=0.5	21,292	4,228,613
Final sample:		4,228,613
Sample with age<=65:		3,326,113

<sup>1</sup>Reasons for a lack of a match are either that the household did not live in an MSA that was identified in the IPUMS data or the identified MSA could not be matched to the MSAs contained both in the REIS and HPI data. <sup>2</sup>MSAs with incomplete HPI data are Greenville, SC and New Haven, CT.