DYNAMIC ASSET PRICING IN A SYSTEM OF LOCAL HOUSING MARKETS

Patrick Bayer, Bryan Ellickson and Paul Ellickson

October 27, 2010

For most people, buying a house is a major investment decision. Economists have mainly focused on the consumption aspects of this process. For example, the purchase of a house might be framed as a discrete choice over a bundle of housing and neighborhood attributes such as location, square footage, schooling options, and crime levels. The investment side of the problem has received considerably less attention, a surprising omission since housing assets comprise approximately two-thirds of the average American household's financial portfolio, serve an important role in saving for retirement and, as has become increasingly apparent, can be quite risky.

While housing purchases are clearly investments, the relevance of modern financial asset-pricing theory to these investments is much less obvious. Most transactions take place between individual owner occupants. From the point of view of an individual homeowner, the market for houses can seem very illiquid and the scope for arbitrage quite limited. However, the owner occupant is often a relatively minor stake-holder in the house that he owns and occupies.

Most homeowners finance a large part of the purchase price of a house by taking out a mortgage, a *derivative* asset whose value depends on (is "derived from") the value of the house. Mortgages are created by commercial banks and other financial institutions. The risk these institutions face that distinguishes a mortgage from a risk-free asset is the risk that the homeowner will default on his mortgage payments, which depends in part on whether the value of the house has fallen below the amount owed on the mortgage. Many mortgages are insured by Fannie Mae and Freddie Mac, which transfers risk from the bank to the insurer. There are other ways to deal with this risk. Banks that issue mortgages can transfer their claim to the payment stream to other financial institutions that pool these claims into *mortgage-based securities*, derivative assets whose value depends on the revenue generated by the payments homeowners choose to make on the mortgages in the pool. Mortgaged-backed securities in turn serve as the underlying asset for derivatives representing claims on various parts (*tranches*) of the revenue stream generated by the mortgage-backed securities, some tranches offering a relatively safe stream of income (analogous to a highly-rated corporate bond) and other tranches bearing the brunt of default by homeowners on their mortgage payments (analogous to corporate stock). Credit default swaps offer another layer of derivatives on top of these derivatives, offering insurance to the holders of highly-rated tranches seeking even stronger protection against the risk of default and, ironically, a vehicle for investors wanting to "short" an inflated housing market.[1] Thus, an individual homeowner who takes out a mortgage to finance 80 or 90 percent of the purchase price of his home is a relatively minor stake-holder in his home. The major stake-holders of owner-occupied houses are, more often than not, highly sophisticated investors holding partial claims to many housing assets, indirect claims in the form of derivatives.

Comparing the financing of a house and a corporate firm is instructive. Consider a corporate firm that has issued publicly-traded common shares and bonds. The stock and bond can themselves be viewed as derivative assets, with the value of the firm serving as the underlying asset-price process.[2] The stock is equivalent to a call option on the value of the firm: if stock-holders exercise the option by paying off the debt, their payoff is the value of the firm minus the payment on the debt. If the prospective payoff is negative, the stock-holders may choose to default. The risk-premium earned by the bond reflects this risk of default. A homeowner is in a position analogous to the stock-holders of a firm. The value of his equity in the home is the value of his option to own the house free and clear if he pays off the debt. If his house is "under water," he may rationally choose to default.

The web of institutions and derivative instruments involved in housing finance is very intricate, but from the point of view of modern finance this complexity is business as usual. Over the past four decades the financial industry has introduced an astonishing array of financial products intended to improve the management of risk. In the case of housing, the benefits from financial innovation seem rather compelling. Until a few years ago, housing market risk was mostly local, with

[1] See Lewis (2010) for an engrossing account of how credit default swaps provided a way for a few investors to bet on a housing market collapse.

[2] See Black and Scholes (1973) and Merton (1974).

the ups and downs of one metropolitan housing market largely uncorrelated with another. However, the originators of mortgages were not so long ago also mostly local institutions, so local risk mattered to them. Mortgage-backed securities and other financial instruments offered a way to pool that risk, allowing local institutions to swap payment streams exposed to local risk for guaranteed payments, enabling them to operate with higher leverage because their debt was (supposedly) safer.

Of course, with the benefit of hindsight we know this story does not end well. Housing risk became increasing national rather than local. Sophisticated financial institutions, failing to recognize how much had changed, mis-priced the risk. Easy credit in turn reinforced the correlation of asset returns in local markets, and eventually the bubble burst.

This paper views housing markets from an asset-pricing perspective, using finance theory to relate the risk premium of a housing asset (the difference between its expected return and the return to a risk-free investment) to its exposure to risk. Thus, we assume that financial institutions were sophisticated enough to eliminate all free lunches. If arbitrage opportunities are eliminated, then all that matters for the risk premium of a housing asset is its exposure to systematic risk. In our model, there are three forms of systematic risk to which housing assets are exposed: *national risk* (which is common to houses everywhere), *metropolitan risk* (which affects all houses within a given metropolitan area, but nowhere else) and *specific risk* (which is systematic risk that is specific to a house type). Houses are said to be of the same type h if they are located in the same metropolitan area and have the same exposure to systematic risk. We assume houses of every type face a common set of risk prices (λ^* for the national risk, λ^m for the local risk specific to metropolitan area m and λ^h for the risk specific to a house of type h) that, together with appropriate measures of exposure to risk, account for the variation in risk-premiums across housing types.

As will be readily apparent, this is a work in progress. Our ultimate goal is to develop an empirical analysis that incorporates most of the metropolitan areas in the United States. Here we focus on 12 local housing markets, 9 located in the four states (Arizona, California, Florida and Nevada) that experienced the worst of the bursting bubble and 3 others that offer some contrast. Within each metropolitan area we define 4 housing types, classified by value. Eventually we plan to estimate all of the parameters of our theoretical model, but here we do not. Instead we simply look at correlations, contrasting the correlations of monthly asset returns between pairs of housing types in two epochs: 1993-2001 (before the bubble) and 2002-2009 (the period containing the bubble and its collapse).

As we will see, the contrast between the two epochs is stark: in the first epoch, most correlations for types within the same metropolitan areas were relatively strong, but almost all correlations for types in different metropolitan areas were not significantly different from 0. In the second epoch, correlations for housing types within metropolitan areas are higher. But the most striking finding is that almost all correlations between housing types belonging to different metropolitan areas are significantly different from zero, and the correlations are often quite high. In the second epoch, housing market risk became national.

Section 1 develops the theoretical model of asset-pricing for a system of local housing markets. Section 2 presents a special case, a model with constant coefficients (which we apply separately to each epoch). We also how the price of a house can be connected to fundamentals, the discounted net rents that it generates. Section 3 explains how we construct the analog of a Case-Shiller index for housing types within a metropolitan area. Section 4 addresses the issue of estimating the covariation and correlation parameters of the model. Section 5 presents the results on the correlations of returns across housing types. Section 6 concludes.

1 A model of housing market risk

The setting is a collection of N single-family housing units located in M metropolitan areas. Besides metropolitan location, houses are classified into K categories. We refer to a specific pairing $h = (m, k)$ as a *housing type*, for example a large house in Los Angeles. The model is formulated as a system of stochastic differential equations (SDE's) driven by a multi-dimensional Wiener process, using as a framework the standard "multidimensional market model" in finance (see Duffie (2001) or Shreve (2004)).

To simplify the exposition, we assume that the local housing markets in our data set are observed over the same interval of time $[0, T] \subset \mathbb{R}$, for example the 20-year period $[0, 20]$. The price process $\widehat{V}^i = (\widehat{V}^i_t)_{t\in[0,T]}$ of house i of type $h = (m, k)$ is assumed to be the solution to the SDE

$$d\widehat{V}^i_t = \widehat{V}^i_t \left[\mu^h_t dt + \sigma^h_t dB^i_t\right] \tag{1}$$

Equation (1) expresses the instantaneous rate of appreciation $d\widehat{V}^i_t / \widehat{V}^i_t$ at time t as the sum of an expected rate of price appreciation $\mu^h_t dt$ and a random shock $\sigma^h_t dB^i_t$, where μ^h_t (the *drift*) and σ^h_t (the *volatility*) are themselves stochastic processes that vary randomly over time. To insure that a solution to equation (1) exists, the processes μ^h and σ^h are assumed to satisfy the conditions[3]

$$\int_0^T |\mu_t^h| dt < \infty \qquad \text{and} \qquad \int_0^T (\sigma_t^h)^2 dB_t^i < \infty$$

$\mathbb{P}$-a.s. (almost surely). The stochastic differential dB_t^i is in turn assumed to be a linear combination of four underlying *risk factors*,

$$dB_t^i = \frac{\sigma_t^{h*}}{\sigma_t^h} dW_t^* + \frac{\sigma_t^{hm}}{\sigma_t^h} dW_t^m + \frac{\sigma_t^{hh}}{\sigma_t^h} dW_t^h + \frac{\sigma_t^{hi}}{\sigma_t^h} dW_t^i \tag{2}$$

where dW_t^*, dW_t^m, dW_t^h and dW_t^i are stochastic differentials of Wiener processes representing *national risk* W^*, *metropolitan risk* W^m, *specific risk* W^h, and *idiosyncratic risk* W^i specific to housing asset i. The coefficients σ_t^{h*}, σ_t^{hm}, σ_t^{hh} and σ_t^{hi} are *covariation parameters* that measure the sensitivity of dB_t^i to the national risk factor, the metropolitan risk factor for metropolitan area m, the risk specific to house type h and the idiosyncratic risk factor specific to house i. The volatility parameter σ^h of equation (1) is linked to the covariation parameters of equation (2) by the following identity,

$$(\sigma_t^h)^2 = (\sigma_t^{h*})^2 + (\sigma_t^{hm})^2 + (\sigma_t^{hh})^2 + (\sigma_t^{hi})^2 \tag{3}$$

The price process of every house is assumed to be governed by a SDE of the form given by equations (1)–(3), yielding a system of SDE's driven by an $M + MK + N + 1$-dimensional Wiener process

$$W = (W^*, W^{m_1}, \dots, W^{m_M}, W^{h_1}, \dots, W^{h_{MK}}, W^{i_1}, \dots, W^{i_N})$$

adapted to a stochastic basis $(\Omega, \mathfrak{F}, \mathbb{F}, \mathbb{P})$.[4]

Finance imposes equilibrium restrictions on this collection of asset-price processes not by relating asset prices to underlying fundamentals but instead by imposing the hypothesis that in equilibrium every possible opportunity for arbitrage has been eliminated. The theory that emerges is a theory of *relative pricing*. The

[3] A process $\widehat{V}^i$ satisfying these requirements is called an Itô process.

[4] The triple $(\Omega, \mathfrak{F}, \mathbb{P})$ is a probability space, consisting of a sample space Ω, a sigma-algebra $\mathfrak{F}$ of subsets of Ω, and a probability measure $\mathbb{P}$ defined on the measurable space $(\Omega, \mathfrak{F})$. The collection $\mathbb{F} = (\mathfrak{F}_t)_{t \in [0,T]}$ is a *filtration*: each $\mathfrak{F}_t$ is a subset of $\mathfrak{F}$, and the sub-sigma algebras $\mathfrak{F}_t$ are increasing (i.e., $\mathfrak{F}_s \subset \mathfrak{F}_t$ if $s \leq t$). We assume $\mathfrak{F} = \mathfrak{F}_1$. By definition, the components of the Wiener process W are independent Wiener processes.

usual choice for a *numéraire* asset is a *bank account process*, an asset price process $B = (B_t)_{t\in[0,T]}$ with initial price $B_0 = 1$ that grows at the risk-free rate r_t: i.e.,

$$B_t = e^{\int_0^t r_s ds} \qquad t \in [0,T] \tag{4}$$

where $r = (r_t)_{t\in[0,1]}$ is a stochastic process satisfying the condition $\int_0^T |r_s| < \infty$ ($\mathbb{P}$ a.s.). The price at time t of housing asset i relative to the bank account is defined as $V_t^i = \widehat{V}_t^i / B_t$. With the relative price V_t^i in place of $\widehat{V}_t^i$, the SDE (1) becomes

$$dV_t^i = V_t^i \left[\alpha_t^h dt + \sigma_t^h dB_t^i\right] \tag{5}$$

where $\alpha_t^h := \mu_t^h - r_t$, the expected rate of price appreciation at time t *net of the risk-free rate*.[5]

When applied in this setting, the theory of asset pricing based on arbitrage asserts that no self-financing portfolio comprised of houses and the risk-free asset can make a positive profit with no risk of loss unless the initial investment in the portfolio is strictly positive $\mathbb{P}$-a.s. The discounted *gain process* $G^i = (G_t^i)_{t\in[0,T]}$ associated with housing asset i is defined by $G_t^i = V_t^i + D_t^i$ for $t \in [0,T]$ where $D_t^i := \int_0^t \rho_t^i dt$ and ρ_t^i is the cash flow (net of expenses and discounted by the bank-account process), received by the owner of the asset at time t. Thus $G_t^i - G_0^i$ is the sum of the *capital gain* $V_t^i - V_0^i$ and the *accumulated net cash flow* D_t^i accruing to an investor holding the asset over the interval $[0,t]$. For a landlord, ρ_t^i is simply the discounted flow of rental income less expenses for maintenance, repairs and the like, which we will refer to as *net rental flow*. For a homeowner, ρ_t^i is the imputed net rental flow.[6]

The *Fundamental Theorem of Asset Pricing* asserts that, provided the housing market eliminates all arbitrage opportunities, there exists a pricing process $Z = (Z_t)_{t\in[0,T]}$ such that the *risk-adjusted* gain process ZG^i for every housing asset i is a martingale: i.e., for all $s,t \in [0,T]$ such that $t \geq s$

$$E(Z_t^i G_t^i \mid \mathfrak{F}_s) = Z_s^i G_s^i$$

where $\mathfrak{F}_s$ is the information set at time s for the stochastic basis $(\Omega, \mathfrak{F}, \mathbb{F}, P)$ on which all of the stochastic processes in the model are defined.

When the price processes are as specified in equations (2), (3) and (5), then the risk-pricing process Z takes a simple form. It is the stochastic process generated

[5] In the finance literature, the relative asset price V_t^i is also called the *discounted* asset price.

[6] Estimating ρ_t^i for homeowners is more difficult than for landlords. As we will see, no-arbitrage theory provides a way around this problem.

by the SDE

$$dZ_t = -Z_t \left[\lambda_t^* dW_t^* + \sum_m \lambda_t^m dW_t^m + \sum_h \lambda_t^h dW_t^h + \sum_i \lambda_t^i dW_t^i \right]$$

with $Z_0 = 1$ where λ_t^*, λ_t^m, λ_t^h and λ_t^i are the market prices of risk at time t associated with the national risk factor W^*, the metropolitan risk factor W^m, the specific risk factor λ^h and the idiosyncratic risk factor W^i respectively. Because by definition idiosyncratic risk has a zero price, $\lambda_t^i = 0$ for all i and this equation simplifies to

$$dZ_t = -Z_t \left[\lambda_t^* dW_t^* + \sum_m \lambda_t^m dW_t^m + \sum_h \lambda_t^h dW_t^h \right] \tag{6}$$

The fact that ZG^i is a martingale implies that this process, which itself is generated by a SDE, must have zero drift. Assume $\rho_t^i / V_t^i = \delta_t^h$ for all houses i of type h. We call δ_t^h the *net rental yield* of houses of type h at time t. Then

$$\alpha_t^h + \delta_t^h = \lambda_t^* \sigma_t^{h*} + \lambda_t^m \sigma_t^{hm} + \lambda_t^h \sigma_t^{hh} \tag{7}$$

for every housing type h.[7] In the finance literature, equations (7), one for each asset type h, are called the *market-price-of-risk equations*. The left-hand side of equation (7) is the *risk premium* of housing type h, the expected instantaneous total return (i.e., capital gains plus net rental yield) at time t, net of the risk-free rate. The right-hand side is the total value of risk exposure for a housing asset of type h, the price of risk times the quantity of risk summed over the three types of systematic risk.

Rather than multiplying the gain G_t^i by Z_t, there is an equivalent way to adjust for risk by changing the probability measure. The value Z_T of the pricing process at time T is a Radon-Nikodym derivative $d\widetilde{\mathbb{P}}/d\mathbb{P}$ that changes the true probability measure $\mathbb{P}$ to an equivalent martingale measure (EMM). Under the EMM $\widetilde{\mathbb{P}}$ the gain process G^i itself, rather than the risk-adjusted gain process ZG^i, is a martingale: i.e., for all $s, t \in [0, T]$ such that $t \geq s$

$$\widetilde{\mathbb{E}}(G_t^i \mid \mathfrak{F}_s) = G_s^i$$

where the tilde on the expectation sign indicates that the conditional expectation is taken with respect to $\widetilde{\mathbb{P}}$. Under the EMM every housing asset generates the

[7]See Shreve (2004).

same gain as the risk-free asset. For this reason, $\widetilde{\mathbb{P}}$ is often called the *risk-neutral probability measure*: by replacing the "true" probability measure by a probability measure that compensates for risk-aversion, assets can be priced "as though" investors are risk neutral, even though they are not.

2 A special case

The theoretical model presented in Section 1 is very general. To be of much use, we need to make specific assumptions concerning the stochastic processes that govern the evolution of the parameters μ_t^h, σ_t^{h*}, σ_t^{hm}, σ_t^{hh} and σ_t^{hi}. In this section, we move to the opposite extreme, assuming that these parameters are constant throughout the entire period $[0, T]$:

$$(\mu_t^h, \sigma_t^{h*}, \sigma_t^{hm}, \sigma_t^{hh}, \sigma_t^{hi}) = (\mu^h, \sigma^{h*}, \sigma^{hm}, \sigma^{hh}, \sigma^{hi}) \qquad (t \in [0, T])$$

We also assume that the risk-free rate r is constant.

Within this special setting, we show how the model can be reinterpreted in terms of discounted net rental flows. It is easier to establish this link if we assume an infinite horizon, so we replace the time set $[0, T]$ with the time set $[0, \infty)$. We take the discounted net rental processes ρ^i as the primitives of our model, demonstrating below that this is equivalent to a model in which the value processes V^i are the primitives. Assume that under the EMM the value of every housing asset equals the expected value of discounted future rents net of expenses. Suppose the process ρ^i for house i of type h is a geometric Brownian motion generated by the SDE

$$d\rho_t^i = \rho_t^i[(\tilde{\eta}^h - r)dt + \sigma^h d\widetilde{B}_t^i] \tag{8}$$

where $\widetilde{B}_t^i$ is a Wiener process under $\widetilde{\mathbb{P}}$. The stochastic differential $d\widetilde{B}_t^i$ is assumed to be a linear combination of national, metropolitan, specific and idiosyncratic risk factors,

$$d\widetilde{B}_t^i = \frac{\sigma^{h*}}{\sigma^h} d\widetilde{W}_t^* + \frac{\sigma^{hm}}{\sigma^h} d\widetilde{W}_t^m + \frac{\sigma^{hh}}{\sigma^h} d\widetilde{W}_t^h + \frac{\sigma^{hi}}{\sigma^h} d\widetilde{W}_t^i \tag{9}$$

where $\widetilde{W}_t^*$, $\widetilde{W}^m$, $\widetilde{W}^h$ and $\widetilde{W}^i$ are Wiener processes under $\widetilde{\mathbb{P}}$ (compare equations (5) and (2) describing the SDE generating the value process V^i). In equation (8), $\tilde{\eta}^h$ is the drift of net rent under $\widetilde{\mathbb{P}}$ (before deflating by the bank-account process). Assume that $\tilde{\eta}^h < r$ and define the discounted value process V^i by $V_t^i = \widetilde{\mathbb{E}}\left[\int_t^\infty \rho_u^i \, du \mid \mathfrak{F}_t\right]$ for $t \in [0, \infty)$. It follows that $V_t^i = \rho_t^i/(r - \tilde{\eta}^h)$. Thus, under the probability measure $\widetilde{\mathbb{P}}$ the net rental yield of a house of type h is the same for all houses i of type h, and it is time invariant. Because $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are equivalent measures, this relationship also holds under $\mathbb{P}$: i.e.,

$$\frac{\rho_t^i}{V_t^i} = \delta^h = r - \tilde{\eta}^h \qquad (\mathbb{P}\text{-a.s}) \tag{10}$$

for all houses i of type h. Letting $\delta^h = r - \tilde{\eta}^h$ in equation (7) we conclude that

$$\alpha^h = \tilde{\eta}^h - r + \lambda^* \sigma^{h*} + \lambda^m \sigma^{hm} + \lambda^h \sigma^{hh} \tag{11}$$

which offers an alternative perspective on the market-price-of-risk equation. If risk-prices are zero (so investors are in fact risk neutral) then $\alpha^h = \tilde{\eta}^h - r$: price appreciation on houses of type h equals the expected rate of increase of net rent minus the risk-free rate. On the other hand, if risk prices are positive and the covariation parameters are positive, then

$$\alpha^h - (\tilde{\eta}^h - r) = \lambda^* \sigma^{h*} + \lambda^m \sigma^{hm} + \lambda^h \sigma^{hh} > 0$$

House values appreciate at a more rapid rate than $\tilde{\eta}^h - r$ to compensate for the risk.

What happens to the process ρ^i under the true probability measure $\mathbb{P}$? Girsanov's Theorem, used to derive equation (7), implies that

$$d\widetilde{B}_t^i = dB_t^i + \left(\frac{\alpha^h + \delta^h}{\sigma^h} \right) dt \tag{12}$$

Substituting (12) into (8) and using (10) to simplify, we obtain $d\rho_t^i = \rho_t^i[\alpha^h dt + \sigma^h dB_t^i]$: under $\mathbb{P}$ the drift in ρ^i matches the drift in V^i. Because V^i is a scalar multiple of ρ^i, it follows that $dV_t^i = V_t^i[\alpha^h dt + \sigma^h d\widetilde{B}_t^i]$, the special case of equation (5) with time-invariant drift and volatility.

We conclude that, in this special case where net rent follows a geometric Brownian motion, 1) the net rent to value ratio is not only constant for all houses of the same type but also time invariant and 2) the growth rate dV_t^i / V_t^i of house value and the growth rate $d\rho_t^i / \rho_t^i$ of net rents are driven by the same Wiener process. By restricting this infinite horizon model to the interval $[0, T]$, these conclusions carry over immediately to our original finite-horizon model.

3 Hedonic Returns

In contrast to financial assets such as stocks or bonds, housing assets are heterogeneous and trade at very low frequency. However, data on repeat sales can be used to overcome these problems. This is the essential insight that led to the construction of the Case-Shiller housing market indexes. In this section we show how the same methodology can be used to construct indexes for houses of different types within a metropolitan area.

Assume that $[0, T]$ is divided into N intervals $[t_{j-1}, t_j]$, corresponding to months. Let $R^i := \log(V^i_t / V^i_s)$ denote the logarithmic return for a housing asset of type $h = (m, k)$ that sells at time s and again at time t, where the selling times $s, t \in [0, T]$ are assumed rounded to the beginning or end of a month. Define $\tau^i := t - s$, the duration of repeat sale i, and let M^i denote the set of months covered by this repeat sale. Define $\Delta t_j := t_j - t_{j-1}$, the length of month j. Similarly, let $\Delta W^*_{t_j}$, $\Delta W^m_{t_j}$ and $\Delta W^h_{t_j}$ denote the increments over month j of the Wiener processes W^*, W^m and W^h respectively.

Assume that the stochastic processes μ^h and σ^h in equation (1) are constant over monthly intervals: i.e.,

$$\mu^h_t = \mu^h_j \qquad \text{and} \qquad \sigma^h_t = \sigma^h_j \qquad \text{for } t \in [t_{j-1}, t_j]$$

Similarly, assume that the stochastic processes σ^{h*}, σ^{hm} and σ^{hh} in equation (2) are constant over monthly intervals

$$\sigma^{h*}_t = \sigma^{h*}_j \qquad \sigma^{hm}_t = \sigma^{hm}_j \qquad \sigma^{hh}_t = \sigma^{hh}_j \qquad \text{for } t \in [t_{j-1}, t_j]$$

and that σ^{hi}_t is constant over the entire period $[0, T]$:

$$\sigma^{hi}_t = \sigma^{hi} \qquad \text{for } t \in [0, T]$$

Using this approximation , it is easy to show that

$$R^i = \gamma^h \tau^i + \sum_{j \in M^i} r^h_j + \varepsilon^i \tag{13}$$

where $\gamma^h := -(\sigma^{hi})^2/2$, $\varepsilon^i := \sigma^{hi}(W^i_t - W^i_s)$, and

$$\begin{aligned} r^h_j \quad := \quad & [\alpha^h_J - \frac{(\sigma^{h*}_j)^2 + (\sigma^{hm}_j)^2 + (\sigma^{hh}_j)^2}{2}]\Delta t_j \\ & + \sigma^{h*}_j \Delta W^*_{t_j} + \sigma^{hm}_j \Delta W^m_{t_j} + \sigma^{hh}_j \Delta W^h_{t_j} \end{aligned}$$

Let N^h be the set of repeat sales of houses of type $h = (m, k)$ over the time interval $[0, T]$. For $j = 1, 2, \ldots, N$ let $I_{\{j \in M^i\}}$ be an indicator variable that equals 1 if month j is covered by the i^{th} repeat sale and 0 otherwise. In regression form equation (13) becomes

$$R^i = \gamma^h \tau^i + \sum_{j=1}^{N} r_j^h I_{\{j \in M^i\}} + \varepsilon^i \tag{14}$$

The coefficient r_j^h in this regression is the portion of the logarithmic return for month j that is common to all housing assets of type h. We refer to the monthly time series $(r_j^h)_{j=1}^N$ generated by these regressions as the monthly *hedonic log returns* for houses of type h.

Equation (14) bears more than a passing resemblance to the methods used by Karl Case and Robert Shiller (1989) to construct housing price indices. The differencing used to obtain the logarithmic return for the i^{th} repeat sale allows us to control for house-specific fixed effects: the constant term V_0^i specific to repeat sale i drops out of the expression for the logarithmic return. Thus, the *level* of housing prices is allowed to be quite heterogeneous, even for houses of the same type. The homogeneity we impose only requires that monthly log returns (the *increment* to log house prices over a month) for houses of the same type are drawn from the same distribution. Because we see only a single realization, in this regression the realizations of W^*, W^m and W^i are fixed, but there are N^h random variables ε^i, one for each repeat sale. By definition of the Wiener processes W^i, the expectation $E\varepsilon^i = 0$ and the disturbances are independently distributed. Consequently, the parameters of equation (14) can be consistently estimated using OLS. Equation (14) highlights two effects of duration on the return. First, the variance of the disturbance term for repeat sale i is $(\sigma^{hi})^2 \tau^i$. As in Case and Shiller, this heteroskedasticity is easily handled. Second, duration has a direct effect on the mean return: the regression coefficient γ^h on the duration τ^i of the i^{th} repeat sale provides an estimate of $-(\sigma^{hi})^2/2$ and hence an estimate for σ^{hi}, the volatility of the idiosyncratic risk for a housing asset i of type h. In this way, deriving equation (14) from a continuous-time structural model leads to a potentially important modification to the classic Case-Shiller specification, a mean correction for duration.

The sum

$$\log \widehat{V}_j^h = r_1^h + r_2^h + \ldots + r_j^h \qquad (j = 1, 2, \ldots, N) \tag{15}$$

defines the cumulative return from holding a house of type h from $t = 0$ until the end of month j. When exponentiated, this yields a price index for houses of type h in a metropolitan area analogous to the Case-Shiller index for the metropolitan area as a whole:

$$\widehat{V}_j^h = 100 \exp(r_1^h + r_2^h + \ldots + r_j^h) \qquad (j = 1, 2, \ldots, N) \tag{16}$$

(as for the Case-Shiller indexes, the initial value $\widehat{V}_0^h$ of the price index is arbitrarily set to 100).

4 Estimating covariation and correlation

The notation $\widehat{V}^h$ is meant to suggest a connection between the price index constructed in Section 3 and the price process $\widehat{V}^i$ for housing asset i discussed in Section 1, but there is an important difference. Because $\widehat{V}_t^i$ is the price of an individual house, the stochastic differential equation (1) governing its evolution depends on idiosyncratic as well as systematic risk, as shown by equation (2). In contrast, the regression that creates the index $\widehat{V}^h$ for houses of type h purges the price index of idiosyncratic risk. Therefore, we assume the stochastic process $\widehat{V}^h$ (viewed as a continuous-time process on $[0, T]$) is the solution to

$$d\widehat{V}_t^h = \widehat{V}_t^i \left[\mu_t^h dt + \sigma_t^h dB_t^h \right] \tag{17}$$

where

$$dB_t^h = \frac{\sigma_t^{h*}}{\sigma_t^h} dW_t^* + \frac{\sigma_t^{hm}}{\sigma_t^h} dW_t^m + \frac{\sigma_t^{hh}}{\sigma_t^h} dW_t^h \tag{18}$$

and

$$(\sigma_t^h)^2 = (\sigma_t^{h*})^2 + (\sigma_t^{hm})^2 + (\sigma_t^{hh})^2 \tag{19}$$

As noted in the introduction, it is instructive to draw a parallel between what we are doing here and the famous Black-Scholes (1973) and Merton (1974) model of the corporate firm. In the Black-Scholes-Merton (BSM) model of the firm, the value of the firm is treated as the fundamental underlying asset. The stock and bond issued by the firm are derivative assets with payoffs that depend on the underlying value of the firm. The BSM model had a large impact on modern finance, but it has a weakness: a corporate firm is rarely sold, so there is little opportunity to observe its underlying value. Housing assets are different. If we are willing to entertain the hypothesis that houses can be classified into a relatively few categories and that houses of the same type have the same exposure to systematic risk, then it is as though in the BSM model we had many copies of the "same" firm, each selling rarely but collectively experiencing a large number of repeat sales.

We now examine the covariation parameters of the model in more detail. Let $\widehat{X}_t^h = \log(\widehat{V}_t^h)$ and let $h = (m, k)$ and $h' = (m', k')$ denote two house types (not necessarily distinct). Using Itô's formula, equation (17) implies that

$$d\widehat{X}_t^h = [\mu_t^h - (\sigma_t^h)^2/2]dt + \sigma_t^h dB_t^h \tag{20}$$

Consequently, the *covariation process* $[\widehat{X}^h, \widehat{X}^{h'}] = ([\widehat{X}^h, \widehat{X}^{h'}]_t)_{t\in[0,T]}$ associated with a pair of housing types $h = (m, k)$ and $h' = (m', k')$ has stochastic differential

$$d[\widehat{X}^h, \widehat{X}^{h'}]_t = \sigma_t^h \sigma_t^{h'} d[B^h, B^h]_t \tag{21}$$

From equation (17) and Itô's rules $d[W^*, W^*]_t = dt$ and

$$d[W^*, W^m]_t = d[W^*, W^h]_t = d[W^m, W^h]_t = 0$$

equation (21) reduces to

$$d[\widehat{X}^h, \widehat{X}^{h'}]_t = \sigma_t^{h*}\sigma_t^{h'*}dt + \sigma_t^{hm}\sigma_t^{h'm'}d[W^m, W^{m'}]_t + \sigma_t^{hh}\sigma_t^{h'h'}d[W^h, W^{h'}]_t$$

The *correlation coefficient* between types h and h' at time t is

$$\rho_t^{hh'} = \frac{d[\widehat{X}^h, \widehat{X}^{h'}]_t}{\sigma_t^h \sigma_t^{h'} dt} \tag{22}$$

We consider three cases:

- *h and h' belong to distinct metropolitan areas: i.e., $m \neq m'$ and hence $h \neq h'$.* Itô's rules imply

  $$d[W^m, W^{m'}]_t = d[W^h, W^{h'}]_t = 0$$

  Equation (21) reduces to

  $$d[\widehat{X}^h, \widehat{X}^{h'}]_t = \sigma_t^{h*}\sigma_t^{h'*}dt \tag{23}$$

  and equation (22) becomes

  $$\rho_t^{hh'} = \frac{\sigma_t^{h*}\sigma_t^{h'*}}{\sigma_t^h \sigma_t^{h'}} \tag{24}$$

  Covariation and correlation depend only on national risk because the other risk dimensions are orthogonal.

- *h and h' belong to the same metropolitan area but the types are distinct: i.e., $m = m'$ but $k \neq k'$.* Itô's rules imply

  $$d[W^m, W^{m'}]_t = dt \qquad \text{and} \qquad d[W^h, W^{h'}]_t = 0$$

  Equation (21) reduces to

  $$d[\widehat{X}^h, \widehat{X}^{h'}]_t = (\sigma_t^{h*}\sigma_t^{h'*} + \sigma_t^{hm}\sigma_t^{h'm})dt \tag{25}$$

  and equation (22) becomes

  $$\rho_t^{hh'} = \frac{\sigma_t^{h*}\sigma_t^{h'*} + \sigma_t^{hm}\sigma_t^{h'm}}{\sigma_t^h \sigma_t^{h'}} \tag{26}$$

  Covariation and correlation depend only on national and metropolitan risk because the specific risk dimensions are orthogonal.

- *The housing types are the same: $h = h'$.* Itô's rules imply

  $$d[W^m, W^{m'}]_t = d[W^h, W^{h'}]_t = dt$$

  Equation (21) reduces to

  $$d[\widehat{X}^h, \widehat{X}^h]_t = [(\sigma_t^{h*})^2 + (\sigma_t^{hm})^2 + (\sigma_t^{hh})^2]dt = (\sigma_t^h)^2 dt \tag{27}$$

  and equation (22) becomes

  $$\rho_t^{hh} = 1 \tag{28}$$

  The covariation of housing type h within itself is the same as the *quadratic variation* $[\widehat{X}^h, \widehat{X}^h]$ of the process $\widehat{X}^h$.

For empirical purposes we need to assume that the quadratic variation parameters σ_t^h and covariation parameters σ_t^{h*}, σ_t^{hm} and σ_t^{hh} are approximately constant over some non-negligible period of time. Mykland and Zhang (2009) have recently developed an elegant large-sample theory to support this common practice. The essential idea is to interpret "local-constancy approximations" of this sort as discrete-time stochastic processes that are *contiguous* to the underlying continuous-time processes.

To define the discrete-time stochastic approximation, Mykland and Zhang employ two grids, finite collections of times that partition the interval $[0, T]$ into subintervals. In our setting, the first grid

$$\mathcal{G} = \{t_0, t_1, \ldots, t_N\}$$

where

$$0 = t_0 < t_1 < \ldots < t_N = T$$

subdivides the interval $[0, T]$ into monthly subintervals $[t_{j-1}, t_j]$, and a first-stage discrete-time approximation on $\mathcal{G}$ is defined by assuming that

$$\widehat{X}^h_t = \widehat{X}^h_{t_{j-1}} \qquad (t \in [t_{j-1}, t_j])$$

We have already employed this approximation in Section 3 when we constructed the housing index $\widehat{V}^h$ for houses of type h, where we assumed for example that the covariation parameter $\sigma^{h*}_t = \sigma^{h*}_{t_{j-1}}$ for all t within month j.[8] We assume that the grid $\mathcal{G}$ is uniformly spaced, setting $\Delta t_j = \frac{1}{12}$: i.e., every month has the same length, one-twelfth of a year.

The second grid $\mathcal{H}$ is a subset of $\mathcal{G}$ which groups the times $t_j \in \mathcal{G}$ into K blocks:

$$\mathcal{H} = \{\tau_0, \tau_1, \ldots, \tau_K\}$$

where

$$0 = \tau_0 < \tau_1 < \ldots < \tau_K = 1$$

The second-level approximation assumes that the first-level approximation is constant over each block: i.e.,

$$\widehat{X}^h_{t_j} = \widehat{X}^h_{\tau_{i-1}} \qquad (t_j \in [\tau_{i-1}, \tau_i])$$

If the covariation parameters of the Itô process $\widehat{X}^h$ are assumed constant over the block $[\tau_{i-1}, \tau_i]$, then the covariation parameters can be estimated by using a scaled version of the corresponding sample covariance of the monthly returns contained within the block. Consider first the quadratic variation process $[\widehat{X}^h, \widehat{X}^h]$ with stochastic differential (27). Rewrite equation (27) as

$$d[\widehat{X}^h, \widehat{X}^h]_t = [\widehat{X}^h, \widehat{X}^h]'_t dt$$

where $[\widehat{X}^h, \widehat{X}^h]'_t = (\sigma^h_t)^2$. The assumption that the quadratic variation parameter σ^h_t is constant over the block $[\tau_{i-1}, \tau_i]$ means that the "slope" $[\widehat{X}^h, \widehat{X}^h]'_t$ of the quadratic variation process is constant over the block:

$$[\widehat{X}^h, \widehat{X}^h]'_t = (\sigma^h_{\tau_{i-1}})^2$$

[8] In Section 3 we wrote σ^{h*}_j for $\sigma^{h*}_{t_{j-1}}$, and used the same shorthand for the other covariation parameters.

If this slope is constant, then the $\widehat{X}^h$ process is (conditional on $\sigma^h_{\tau_{i-1}}$) a Brownian motion over the block, which in turn implies that the process sampled once a month has i.i.d. returns that are normally distributed. Letting M_i denote the number of months $[t_{j-1}, t_j]$ contained in the block $[\tau_{i-1}, \tau_i]$, the uniform minimum variance unbiased (UMVU) estimator of the slope is

$$\widehat{(\sigma^h_t)^2} = \frac{1}{\Delta t_j (M_i - 1)} \sum_{t_j \in (\tau_{i-1}, \tau_i]} (\Delta X^h_{t_j} - \overline{\Delta X^h_{\tau_{i-1}}})^2 \tag{29}$$

where

$$\overline{\Delta X^h_{\tau_{i-1}}} = \frac{1}{M_i} \sum_{t_j \in (\tau_{i-1}, \tau_i]} \Delta X^h_{t_j} = \frac{1}{M_i} (X^h_{\tau_i} - X^h_{\tau_{i-1}}) \tag{30}$$

Equation (29) corresponds to equation (54) in Mykland and Zhang (2009). If Δt_j were not present, equation (29) would be the usual UMVU estimator of the variance of M_i i.i.d. normally distributed random variables. The time increment $\Delta t_j = 1/12$ scales this estimator of the variance so that it becomes an annualized rate of change of the quadratic variation process $[\widehat{X}^h, \widehat{X}^h]$.

UMVU block estimates for the slopes $[\widehat{X}^h, \widehat{X}^h]'_t$ for $h \neq h'$ are obtained in the same way. The stochastic differential (23) and the stochastic differential (25) can be written in the form

$$[\widehat{X}^h, \widehat{X}^h]_t = [\widehat{X}^h, \widehat{X}^h]'_t dt$$

where

$$[\widehat{X}^h, \widehat{X}^h]'_t = \sigma^{h*}_t \sigma^{h'*}_t$$

in equation (23) and

$$[\widehat{X}^h, \widehat{X}^h]'_t = \sigma^{h*}_t \sigma^{h'*}_t + \sigma^{hm}_t \sigma^{h'm}_t$$

in equation (25). In either case, the UMVU estimator of the slope equals the scaled sample covariance:

$$\widehat{[\widehat{X}^h, \widehat{X}^h]'_t} = \frac{1}{\Delta t_j (M_i - 1)} \sum_{t_j \in (\tau_{i-1}, \tau_i]} (\Delta X^h_{t_j} - \overline{\Delta X^h_{\tau_{i-1}}})(\Delta X^{h'}_{t_j} - \overline{\Delta X^{h'}_{\tau_{i-1}}}) \tag{31}$$

The block estimator of the corresponding correlation coefficients is simply the sample correlation coefficient for the monthly returns in the block, which requires no scaling.

To obtain large-sample results, Mykland and Zhang consider sequences $(\mathcal{G}^n)_{n\geq 1}$ and $(\mathcal{H}^n)_{n\geq 1}$ of grids that subdivide the interval $[0, T]$ into finer increments Δt_j^n and shorter block lengths $\Delta \tau_i^n$ as $n \to \infty$. These asymptotic results justify the use of block estimators to estimate the increment $[\widehat{X}^h, \widehat{X}^{h'}]_T - [\widehat{X}^h, \widehat{X}^{h'}]_0$ to covariation over the period $[0, T]$. Our focus is the volatility or covariation estimates for each block rather than the estimates of the integrated volatility or covariation over the entire period $[0, T]$. Eventually we intend to use blocks that cover 2 or 3 years of monthly log returns. However, in this version of the paper, we divide the period $[0, T]$ into two large blocks (which we call *epochs*): $[\tau_0, \tau_1]$ extending from 1993 to the end of 2001, and $[\tau_1, \tau_2]$ extending from the beginning of 2002 to the end of 2009.

5 Correlations across types

To recapitulate, we have adapted a standard model of modern finance to housing markets and derived a relationship that relates the reward for holding a housing asset to the risk. We have also established a link between the value of a house, which is only infrequently observed (at the point of sale), and an index of the value of houses of that type — an index that generalizes the Case-Shiller indexes for metropolitan areas to indexes for types of houses within a metropolitan area. However, up to now a housing type has been rather abstract: a housing type $h = (m, k)$ identifies a house by its location in a metropolitan area m and its characteristics k. The time has come to be more specific. We characterize houses by their location in a metropolitan area and by their *relative value*.

Characterizing housing by their relative value follows a long tradition of studying stock markets, where portfolios of stocks ranked by *size* (capitalization, the total value of shares outstanding) have played an important role in testing asset-pricing models from the earliest tests of the CAPM to the present. We have every reason to think that "size matters" for houses as well as stocks — what else was the sub-prime mortgage crisis about?

However, as with so many aspects of this project, houses raise complications. Ranking stocks by value is easy, because stocks trade frequently. Ranking firms by value would be much more difficult, because firms are seldom sold. Houses are an intermediate case, sold more frequently than most firms but less frequently than stock in those firms. Less abstractly, the problem is that we can observe, essentially continuously, the stock price of every publicly-traded firm, but the set of houses put up for sale is endogenous. It is easy to believe that, in the midst of the bubble, exchanges of relatively low-value houses dominated the market, but that — after the crash — the market for low-value houses evaporated.

The results reported in this section take the easy way out. We rank all repeat sales over the period 1993-2009 by the first of the two prices in the repeat sale and classify the repeat sales into quartiles from Quartile 1 for the bottom 25% to Quartile 4 for the top 25%. Obviously we need to do better, and we are working on that. But for now our quartiles are defined only for the entire pooled sample of all repeat sales from 1993 through 2009, with no adjustment for inflation or for the possibility that the pool of homeowners who choose to sell is endogenous.

As noted at the end of Section 4, in this paper we divide the sample interval $[0, T]$ into two blocks (*epochs*), 1993–2001 and 2002–2009. Within each block $[\tau_{i-1}\tau_i]$ we compute sample correlations $\widehat{\rho}^{hh'}_{\tau_{i-1}}$. The odd-numbered tables 1,3,5,7 report the correlations for 1993-2001, and the even-numbered tables 2,4,6,8 report the correlations for 2002-2009. Recall from Section 4 that the correlations for types h, h'' from *different* metropolitan areas depend only on *national* risk:

$$\rho^{hh'}_{\tau_{i-1}} = \frac{\sigma^{h*}_{\tau_{i-1}} \sigma^{h'*}_{\tau_{i-1}}}{\sigma^{h}_{\tau_{i-1}} \sigma^{h'}_{\tau_{i-1}}}$$

Because the other components of systematic risk are orthogonal, they cancel out.

Reading Tables 1–8 is quite straightforward, and the results are striking. Each table represents the part of the sample correlation matrix below the diagonal. For example, Tables 1 show the correlations $\rho^{hh'}_{\tau_0}$ for pairs of quartile 1 houses in epoch $[\tau_0, \tau_1]$ (1993-2001), type h chosen from one metropolitan area and type h' from a different metropolitan area. To enhance readability, correlations are multiplied by 100, and those correlations significant at the 0.01 level are indicated by an asterisk. The columns and rows are labeled by metropolitan area:

LA	Los Angeles	SD	San Deigo	SF	San Francisco
MI	Miami	OR	Orlando	TA	Tampa
LV	Las Vegas	PH	Phoenix	TU	Tucson
CH	Chicago	CL	Cleveland	DN	Denver

The contrast between epochs is rather dramatic. In table 1, only 2 of the correlation coefficients are significantly different from zero. In table 2, all 36 correlations involving metropolitan areas in the four states most affected by the bubble (California, Florida, Arizona and Nevada) are significantly different from 0, positive and large. For the three metropolitan areas we added for contrast (Chicago, Cleveland and Denver), 6 of 9 correlation coefficients involving Chicago are statistically significant, but none involving Cleveland and only one for Denver.

The patterns for the other quartiles are similar. For Quartile 2, two correlation coefficients are significantly different from 0 in epoch 1 (Table 3), but all correlation coefficients in the bubble states are statistically significant, positive and large in epoch 2 (Table 4). Chicago shows evidence of exposure to national risk in epoch 2, but Cleveland and Denver for the most part do not. For Quartile 3, two correlation coefficients are significantly different from 0 in epoch 1 (Table 5), while all correlations in the bubble states are statistically significant, positive and large in epoch 2 (Table 6).

Only for Quartile 4 do we see any change in the pattern. In epoch 1, 8 correlation coefficients are statistically different from 0 (Table 7). Three of these are between metropolitan areas in California and one between Orlando and Tampa in Florida. In epoch 2, most of the 36 correlation coefficients in the bubble states are significantly different from 0, but two are not, and the correlation coefficients tend to be somewhat smaller than was the case for Quartiles 1–3. We expected to see more attenuation of exposure to national risk as we ascended the value ladder, but this is all we see. We believe the failure to see more attenuation is largely due to the lack of rebalancing of our quartiles over time, an issue which is now being addressed.

Looking at correlations rather than covariation has the advantage of being scale free, but it does not take full advantage of our structural model. The 4 tables associated with each epoch present 264 correlation coefficients,

$$\rho_{\tau_{i-1}}^{hh'} = \frac{\sigma_{\tau_{i-1}}^{h*}\sigma_{\tau_{i-1}}^{h'*}}{\sigma_{\tau_{i-1}}^{h}\sigma_{\tau_{i-1}}^{h'}}$$

From the perspective of our theoretical model, our main interest is in the numerator of this correlation coefficient, the product $\sigma_{\tau_{i-1}}^{h*}\sigma_{\tau_{i-1}}^{h'*}$ that is estimated by equation (31), the *scaled sample covariation* of monthly log returns over the block. Two covariation parameters of the structural model ($\sigma_{\tau_{i-1}}^{h*}$ and $\sigma_{\tau_{i-1}}^{h'*}$) enter each product, and there are only 48 of these covariation parameters — 4 for each of our 12 metropolitan areas. Rather than look at correlation coefficients, a more parsimonious and informative procedure is to match scaled sample covariances with the products $\sigma_{\tau_{i-1}}^{h*}\sigma_{\tau_{i-1}}^{h'*}$, using GMM. The system is over identified: 264 moment conditions for 48 structural parameters $\sigma_{\tau_{i-1}}^{h*}$. In fact, this is an undercount because we have considered only pairs $h = (m, k)$ and $h = (m', k')$ for which $m \neq m'$ (the metropolitan areas are different) but $k = k'$ (the value quantile is the same). If we allow $k \neq k'$ as well, then there are 1056 moment conditions to determine the 48 structural parameters $\sigma^{h*}_{\tau_{i-1}}$. Phrased another way, the parameter $\sigma^{h*}_{\tau_{i-1}}$ measuring the exposure of low-value houses in Los Angeles to national risk enters 44 moment conditions, one for each pairing of a low-value houses in LA and one of the 4 house types in the 11 metropolitan areas other than LA. GMM will produce *one* estimate for the LA parameter, using the information contained in 44 matches. These considerations will become even more compelling as we expand our list of metropolitan areas: the structural parameters increase linearly with the number of metropolitan areas, the moment conditions increase with the square.

Tables 9 and 10 exploit a second type of orthogonality condition by examining the correlation between distinct quartiles *within* the same metropolitan area. In equation (26) we derived an expression relating the correlation coefficient to the structural parameters of the theoretical model: for epoch i,

$$\rho^{hh'}_{\tau_{i-1}} = \frac{\sigma^{h*}_{\tau_{i-1}}\sigma^{h'*}_{\tau_{i-1}} + \sigma^{hm}_{\tau_{i-1}}\sigma^{h'm}_{\tau_{i-1}}}{\sigma^{h}_{\tau_{i-1}}\sigma^{h'}_{\tau_{i-1}}}$$

where $h = (m,k)$ and $h' = (m,k')$. Because $k \neq k'$ (the quartiles are distinct), type-specific risk drops out of the numerator of this correlation coefficient.

The columns of Tables 9 and 10 are labeled by metropolitan area. The rows are labeled by quartile pair for each of the 6 possible pairings of distinct quartiles. In epoch 1, 20 of the correlation coefficients are significantly different from 0, 14 in the three California metropolitan areas and two in Florida. In epoch 2, all of the correlation coefficients for cities in California and Florida are statistically significant, as are all of the coefficients for Los Vegas and 5 out of 6 for Tucson, but only 2 of 6 for Phoenix. All six are significant for Chicago and Denver, but none for Cleveland.

To disentangle national from metropolitan risk, we again turn to GMM. The first set of orthogonality conditions should provide reliable estimates of national risk. Equating $\sigma^{h*}_{\tau_{i-1}}\sigma^{h'*}_{\tau_{i-1}} + \sigma^{hm}_{\tau_{i-1}}\sigma^{h'm}_{\tau_{i-1}}$ to the corresponding scaled sample moment for each of the 6 pairings of distinct types h and h' within a metropolitan area gives six moment conditions to determine the metropolitan-risk covariation parameters σ^{hm} for each of the four value quantiles. In contrast to the moment conditions used to determine the national risk parameters, this system is only slightly over identified because we have tacitly assumed no restriction on the relationship between a covariation parameter $\sigma^{hm}_{\tau_{i-1}}$ for Quantile 1 in metropolitan area m (Miami) and the corresponding parameter $\sigma^{h'm}_{\tau_{i-1}}$ for Quantile 1 in metropolitan area m' (Tampa).

It is probably unrealistic to assume that $\sigma^{h'm}_{\tau_{i-1}}$ for a given value quantile is the same in all metropolitan areas, but perhaps useful to impose such restrictions for metropolitan areas sharing features in common (e.g., belonging to the same state).

All that remains to be determined are the covariation parameters $\sigma^{hh}_{\tau_{i-1}}$ measuring sensitivity to type-specific systematic risk. From equation (19) we have

$$(\sigma^{h}_{\tau_{i-1}})^2 = (\sigma^{h*}_{\tau_{i-1}})^2 + (\sigma^{hm}_{\tau_{i-1}})^2 + (\sigma^{hh}_{\tau_{i-1}})^2$$

The parameters $\sigma^{h*}_{\tau_{i-1}}$ are determined by the first set of orthogonality conditions and the parameters $\sigma^{hm}_{\tau_{i-1}}$ by the second. The left-hand side is matched to the scaled sample variance for type h, equation (29). This determines the type specific parameter $\sigma^{h*}_{\tau_{i-1}}$, and the (scaled) variance decomposition is complete.

6 Conclusions

In academic finance, residential real estate is a backwater. In the real world, it has become a major focus for financial innovation, with consequences both good and bad. We are fortunate that this exceptional time of financial turmoil has coincided with the availability of increasingly rich data on housing prices and housing finance. Our goal is to establish a connection between these rich data sets and modern asset-pricing theory.

As we noted at the outset, this is a work in progress. In Section 1 we developed an asset-pricing model for housing prices, modeling these prices as stochastic processes with continuous paths that evolve randomly over time. Section 2 illustrated a special case where the coefficients of these processes remained constant over time. Unlike stocks or bonds, houses trade rather infrequently. In Section 3, we addressed that problem by constructing monthly price series for *housing types* within a metropolitan area, using the same techniques that Case and Shiller (1989) used to construct national and metropolitan housing indexes. Such an approach is likely to strike many readers as inappropriate for asset markets as apparently illiquid as residential real estate. However, the illiquidity of this market depends very much on ones' point of view. For the individual homeowner seeking to sell his home, or the prospective buyer looking for a suitable house, search looms large. But for the institutions that finance home purchases, houses are much closer to commodities. Whether we are on the right track is clearly debatable. The purpose of this paper is to perform a simple test: divide the data we have into two epochs, and test whether our approach can detect an increased sensitivity to national risk in the second epoch. To set up this test, we used as our housing types value quantiles, and examined correlations among these quantiles. Section 4 uses the theoretical model to establish a connection between these correlations and the parameters of our model. If we are on the right track, we should be able to detect an increased exposure to national risk in the second epoch, the period containing the bubble and burst. The evidence strongly suggests that this is the case.

It could be argued that this is no surprise, because prices in metropolitan areas obviously went up and down in coordinated fashion during the boom and bust. However, the correlations we examine are not in levels but in first differences of the price indexes (monthly log returns). The estimation technique we use, patterned after the block estimation technique of Mykland and Zhang (2009), takes out the trend, so the test has some content. However, examining correlations is an indirect test of the model. Covariation is more direct, but the relevant covariation parameters must be unpacked using orthogonality restrictions. In Section 5, we describe the way we propose to conduct that estimation using GMM.

One of the advantages of using GMM is that, because the model is heavily parameterized, we can reduce the size of the blocks we use in our estimates. Assuming that covariation parameters abruptly changed from epoch 1 to epoch 2 is clearly unrealistic. Much more palatable would be an assumption that the parameters changed continuously, though perhaps rapidly as the pace of financial innovation accelerated. We suspect that we can reduce block sizes to perhaps 2–3 years, yielding 24–36 months of log returns per block — perhaps even shorter. If this works, what the estimates would produce is an admittedly somewhat granular time series for the covariance parameters from 1993 through 2009. Coupled with data on what was happening in mortgage markets and a model of how households reacted to these changing conditions, we hope to shed some light on how this episode of financial innovation turned out so badly and how we might learn from the experience.

References

[1] Bayer, Patrick, Bryan Ellickson and Paul B. Ellickson. 2010. "Dynamic Asset Pricing in a System of Local Housing Markets", *American Economic Review Papers and Proceedings* 100(2): 368–372.

[2] Black, Fischer, and Myron Scholes. 1973. "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81(3): 637–654.

[3] Case, Karl E., and Robert J. Shiller. 1989. "The Efficiency of the Market for Single-Family Homes," *American Economic Review* 79(1): 125–137.

[4] Duffie, Darrell. 2001. *Dynamic Asset Pricing Theory (Third Edition).* Princeton University Press, Princeton.

[5] Gorton, Gary B. 2010. *Slapped by the Invisible Hand: The Panic of 2007*. Oxford University Press, Oxford.

[6] Lewis, Michael 2010, *The Big Short: Inside the Doomsday Machine*, Norton, New York.

[7] Merton, Robert C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *Journal of Finance* 29(2): 449–470.

[8] Mykland, Per A. and Lan Zhang 2009. "Inference for Continuous Semimartingales Observed at High Frequency," *Econometrica* 77(5), 1403–1445.

[9] Muth, Richard, 1969. *Cities and Housing*. University of Chicago Press, Chicago.

[10] Shreve, Steven E. 2004. *Stochastic Calculus for Finance: II*. Springer-Verlag, New York.

Table 1: Quartile 1 correlations (×100), 1993–2001

	LA	SD	SF	MI	OR	TA	LV	PH	TU	CH	CL
SD	-2										
SF	23*	12									
MI	26	22	3								
OR	3	6	-14	-5							
TA	-16	-5	1	-14	0						
LV	-13	18	13	8	12	15					
PH	22	5	2	6	15	-15	8				
TU	29*	-5	-14	26	10	-5	-15	11			
CH	8	47*	3	0	18	-27	17	16	-18		
CL	17	-12	6	-26	-10	26	-10	19	-5	-12	
DN	3	12	29	10	12	35	10	-28	16	-9	-4

Table 2: Quartile 1 correlations (×100), 2002–2009

	LA	SD	SF	MI	OR	TA	LV	PH	TU	CH	CL
SD	67*										
SF	75*	63*									
MI	75*	55*	62*								
OR	51*	41*	44*	49*							
TA	66*	49*	50*	58*	50*						
LV	63*	51*	69*	69*	61*	62*					
PH	59*	53*	65*	67*	61*	58*	76*				
TU	39*	30*	28*	29*	37*	45*	39*	44*			
CH	27	28*	43*	25	42*	28*	39*	35*	10		
CL	22	10	9	24	6	-4	-3	21	8	-11	
DN	37*	7	24	3	4	8	9	17	5	15	12

Table 3: Quartile 2 correlations (×100), 1993–2001

	LA	SD	SF	MI	OR	TA	LV	PH	TU	CH	CL
SD	18										
SF	34*	11									
MI	25	16	22								
OR	-1	-25	11	-16							
TA	21	10	-9	-1	7						
LV	1	22*	18	5	-11	-27					
PH	18	0	13	7	-5	-10	10				
TU	-6	-8	7	21	-24	-5	-9	5			
CH	22	1	12	25	-5	9	-9	29	-28		
CL	29	3	17	9	12	19	23	23	0	2	
DN	40*	-11	27	8	-12	-4	-2	-20	12	-2	19

Table 4: Quartile 2 correlations (×100), 2002–2009

	LA	SD	SF	MI	OR	TA	LV	PH	TU	CH	CL
SD	76*										
SF	79*	64*									
MI	67*	56*	64*								
OR	61*	54*	48*	61*							
TA	65*	49*	51*	66*	63*						
LV	79*	63*	72*	69*	56*	54*					
PH	70*	53*	65*	69*	62*	69*	68*				
TU	43*	40*	38*	47*	41*	58*	40*	68*			
CH	28	35*	39*	35*	40*	47*	43*	38*	28*		
CL	12	8	20	30*	7	14	10	9	15	0	
DN	27	29*	42*	28*	9	11	21	21	11	22	24

Table 5: Quartile 3 correlations (×100), 1993–2001

	LA	SD	SF	MI	OR	TA	LV	PH	TU	CH	CL
SD	20										
SF	38*	6									
MI	-5	-5	8								
OR	19	18	0	-18							
TA	34*	12	-5	32	5						
LV	4	4	-5	16	8	29					
PH	22	-8	-2	1	13	0	21				
TU	-6	-4	7	7	-20	19	22	-8			
CH	-8	23	12	8	-6	-10	0	-10	3		
CL	15	-8	8	12	32	8	27	15	-24	-4	
DN	36	-7	15	-30	9	-1	8	9	-9	6	4

Table 6: Quartile 3 correlations (×100), 2002–2009

	LA	SD	SF	MI	OR	TA	LV	PH	TU	CH	CL
SD	70*										
SF	71*	63*									
MI	78*	52*	57*								
OR	68*	37*	55*	72*							
TA	58*	44*	58*	59*	51*						
LV	76*	56*	58*	61*	58*	53*					
PH	65*	52*	63*	79*	75*	62*	59*				
TU	46*	42*	42*	49*	60*	45*	42*	65*			
CH	21	18	26	27*	30*	28*	28*	31*	30*		
CL	23	23	36*	15	20	10	15	24	14	28*	
DN	23	17	38*	20	20	23	23	19	29*	8	27*

Table 7: Quartile 4 correlations (×100), 1993–2001

	LA	SD	SF	MI	OR	TA	LV	PH	TU	CH	CL
SD	26*										
SF	37*	22*									
MI	-22	17	-2								
OR	-2	6	-3	11							
TA	3	20	-21	32	50*						
LV	-7	13	-11	3	19	10					
PH	10	-4	20	3	-11	2	4				
TU	-22	-13	-9	42*	19	35*	31*	-5			
CH	25	7	-1	9	6	8	-3	24	6		
CL	18	10	21	4	14	12	30	5	6	1	
DN	13	14	8	33	-19	42*	4	11	-7	15	33

Table 8: Quartile 4 correlations (×100), 2002–2009

	LA	SD	SF	MI	OR	TA	LV	PH	TU	CH	CL
SD	59*										
SF	42*	41*									
MI	46*	31*	29*								
OR	57*	34*	31*	58*							
TA	51*	39*	36*	42*	64*						
LV	68*	44*	38*	47*	56*	49*					
PH	52*	41*	37*	64*	72*	57*	58*				
TU	20	21	33*	19	52*	44*	35*	43*			
CH	10	4	7	30*	18	1	21	28*	-13		
CL	13	20	40*	1	34*	26	13	14	42*	-14	
DN	23	13	34*	21	27	19	27*	27*	21	14	28*

Table 9: Intra-metropolitan correlations between types (×100), 1993–2001

Pair	LA	SD	SF	MI	OR	TA	LV	PH	TU	CH	CL	DN
1,2	45*	16	32*	6	12	-2	2	3	10	20	3	23
1,3	32*	20*	29*	3	34*	7	-17	-15	10	-12	15	27
1.4	28*	4	16	32	-9	22	-3	-37*	10	11	-19	25
2,3	59*	23*	44*	23	32*	20	3	0	33*	9	27	47*
2,4	47*	17	32*	5	12	9	6	25*	2	22	35*	13
3,4	58*	26*	47*	2	-21	19	8	-14	7	-2	-29	31

Table 10: Intra-metropolitan correlations between types (×100), 2002–2009

Pair	LA	SD	SF	MI	OR	TA	LV	PH	TU	CH	CL	DN
1,2	92*	66*	67*	62*	54*	61*	80*	80*	45*	60*	17	62*
1,3	87*	59*	67*	72*	50*	47*	68*	80*	42*	54*	0	51*
1.4	67*	52*	38*	61*	34*	40*	61*	67	24	27*	7	44*
2,3	90*	68*	71*	69*	46*	43*	83*	87	53*	64*	20	55*
2,4	77*	63*	45*	55*	57*	54*	77*	76	37*	55*	15	42*
3,4	81*	72*	53*	69*	64*	56*	91*	84	29*	53*	26	46*