



Institute of  
Business and  
Economic Research

Fisher Center for  
Real Estate and  
Urban Economics

---

# **PROGRAM ON HOUSING AND URBAN POLICY**

WORKING PAPER SERIES

**WORKING PAPER NO. W08-007**

## **HOUSING PRICE DYNAMICS IN TIME AND SPACE: PREDICTABILITY, LIQUIDITY AND INVESTOR RETURNS**

By

Min Hwang  
John M. Quigley

May 2010

These papers are preliminary in nature: their purpose is to stimulate discussion and comment. Therefore, they are not to be cited or quoted in any publication without the express permission of the author.

---

UNIVERSITY OF CALIFORNIA, BERKELEY

# Housing Price Dynamics in Time and Space: Predictability, Liquidity and Investor Returns

Min Hwang · John M. Quigley

Published online: 14 November 2009

© The Author(s) 2009. This article is published with open access at Springerlink.com

**Abstract** It is widely accepted that aggregate housing prices are predictable, but that excess returns to investors are precluded by the transactions costs of buying and selling property. We examine this issue using a unique data set—all private condominium transactions in Singapore during an eleven-year period. We model directly the price discovery process for individual dwellings. Our empirical results clearly reject a random walk in prices, supporting mean reversion in housing prices and diffusion of innovations over space. We find that, when house prices and aggregate returns are computed from models that erroneously assume a random walk and spatial independence, they are strongly autocorrelated. However, when they are calculated from the appropriate model, predictability in prices and in investment returns is completely absent. We show that this is due to the illiquid nature of housing transactions. We also conduct extensive simulations, over different time horizons and with different investment rules, testing whether better information on housing price dynamics leads to superior investment performance.

**Keywords** Housing market liquidity · Price discovery · Spatial correlation

**JEL Codes** E31 · C23 · R32

---

We thank for comments from Piet Eichholtz, Alan Gelfand, Yuming Fu, Sau Kim Lum and Steve Malpezzi, seminar participants at the 2006 AREUEA meeting in Boston, 2007 Weimer School for Advanced Studies in Real Estate and Land Economics conference and the 2008 Maastricht-MIT-NUS Real Estate Finance and Investment Symposium, and an anonymous referee of this journal. A previous version of this paper was circulated as “Price Discovery in Time and Space: The Course of Condominium Prices in Singapore.”

---

M. Hwang (✉)

George Washington University, Washington, DC, USA

e-mail: min@gwu.edu

J. M. Quigley

University of California, Berkeley, CA, USA

e-mail: quigley@econ.berkeley.edu

## Introduction

The durability, fixity and heterogeneity of dwellings imply that transactions costs are significant in the housing market. In comparison to financial markets, and in comparison to the markets for most consumer goods, housing purchases require costly search to uncover the prices and attributes of commodities. Given the many frictions associated with the purchase of housing, it is hardly surprising that observed price behavior deviates from that predicted by simple models of economic markets.

Inertia in the price adjustment process in the housing market has been well-documented; many studies have concluded that returns in housing markets are predictable by demonstrating that time series estimates of aggregate prices exhibit inertia (Hosios and Pesando 1991; Guntermann and Norrbin 1991; Gatzlaff 1994; and Malpezzi 1999).

Persistence in housing returns has been found recently at the level of individual dwellings—in eight regional markets in Sweden (Englund et al. 1999) and in four U.S. cities (Hill et al. 1999). These analyses reject the random walk hypothesis, which underlies the Case and Shiller model (1989), in favor of a first-order serial correlation process for housing prices. But in this geographical market, price signals exist in space as well as time. Many of the features which can lead to slow diffusion in the time domain may have analogous effects over space. Price information diffuses over space as well as time, and information costs alone can cause prices to deviate from random fluctuations.

Disentangling a diffusion process in the time domain and a diffusion process over space is not straightforward. Our analysis concludes that it is important to take explicit account of these diffusion processes—since shocks and signals tend to be substantially more persistent when individual dwelling returns are correlated over time and over space. Our analysis shows that, due to this persistence, conventional procedures for estimating price series are quite likely to generate misleading estimates of aggregate housing returns.

In this paper, we develop a model of price diffusion, and we incorporate a more general and more appropriate structure of the price discovery process at the level of the individual dwelling. We explicitly incorporate spatial and temporal dependence of idiosyncratic housing prices into a repeat transactions model. We construct a repeat transactions index of prices by estimating this model using a body of data almost uniquely suited to the analysis: the prices of condominium dwellings in Singapore using all transactions reported in the entire country during an eleven-year period. Multiple transactions of the same condominium unit are observed, and all dwellings with market transactions are geocoded. We compare the properties of aggregate housing price indexes and returns computed from our more general model with indexes computed from conventional models. We find that when aggregate investment returns are estimated from models permitting housing prices to follow a random walk and that they be spatially independent, aggregate housing returns are strongly predictable. However, when aggregate returns are estimated from more general models permitting mean revision and spatial correlation, predictability in aggregate investment returns is completely absent. We show that this arises from two sources; the illiquid nature of housing transactions and the persistent shocks to

individual housing prices in time and over space. With these two factors present, the estimated predictability in aggregate housing returns is biased. In principle, even with persistent shocks on individual housing prices, if there were enough transactions at a given point in time, then this small sample bias would disappear. However, we also find that having more observations over time, while maintaining a constant transaction frequency, exacerbates this bias. Only higher transactions frequencies would reduce the bias. For the bias to disappear, however, the frequency of transactions would need to be about as high in the housing market as it is in the stock market.

We then analyze the economic implications of these statistical findings for investment in housing markets. In particular, we simulate the investment outcomes for an investor fully informed about spatial and temporal dynamics with the outcomes for an uninformed investor. Presumably, better information about housing market dynamics will lead to better investment performance in the housing market. We find that the investor with better knowledge of price diffusion over time and space outperforms the uninformed investor, capitalizing on this informational advantage. However, her superior performance appears to be bounded by relatively short holding periods and low transactions costs.

“[A Micro Model of House Prices](#)” develops a general model of housing prices that supports explicit tests for the spatial and temporal pattern of price movements. This section links our model to the widely employed method for measuring housing prices proposed more than 40 years ago by Bailey et al. (1963), as well as its subsequent extensions (e.g., Case and Shiller 1987). The data are described tersely in “[Data](#)”. Our empirical results are presented in “[The Diffusion of House Price Innovations](#)”, “[The Course of Housing Prices, Investment Returns and Their Predictability](#)”, “[Investment Performance](#)” and “[Conclusion](#)”. We test for random walks in space and time against the alternative of mean reversion, and we examine the link between pricing deviations at the individual level and aggregate price movements. We reconcile the puzzle of autocorrelated estimates of prices and returns. We also investigate investor behavior and housing market illiquidity in some detail.

## **A Micro Model of House Prices**

The objects of exchange in the housing market are imperfect substitutes for one another. Indeed, dwellings with identical physical attributes may differ in market price simply because the price incorporates a complex set of site-specific amenities and access costs. But few dwellings have identical physical characteristics; thus comparison-shopping is more difficult and more expensive than in most other markets.

Moreover, housing transactions are made only infrequently, so households must consciously invest in information to participate in this market. As a result, the market is characterized by a costly matching process. Market agents, buyers, and sellers are heterogeneous, and they differ in information and motivation; commodities are themselves heterogeneous. Consequently an observed transaction price for a specific unit may deviate from the price ordained in a simpler environment.

Buyers, sellers, appraisers, and real estate agents estimate the “market price” of a dwelling by utilizing the information embodied in the set of previously sold dwellings. The usefulness of these transactions as a reference depends upon their similarity across several dimensions: physical, spatial, and temporal. Inferences about the “market price” of the dwelling can be drawn only imperfectly from a set of past transactions, because dwellings differ structurally, enjoy different locational attributes, and are valued under different market conditions by different actors over time. Because dwellings trade infrequently, the arrival of new information about market values is slow. From an informational standpoint, the closest comparable transaction across these various dimensions may be the last transaction of the same dwelling. Alternatively, the most comparable transaction may be the contemporaneous selling price of another dwelling in close physical proximity.

An attempt to uncover the market value of a dwelling is further complicated by the fact that a transactions price is not only a function of observable physical characteristics, but also of unobserved buyer and seller characteristics such as their urgency to conclude a transaction (Quan and Quigley 1991). For any given transaction, all that is known is that an offer was made by a specific buyer that was higher than a specific seller’s reservation price.

We develop a model with spatially and temporally correlated errors in a repeat transactions framework. Innovations in housing prices are assumed to take place continually, but transactions are observed sporadically. At any point in time, the prices of houses are dependent over space. In the determination of the price of a house, the weights attributable to neighboring houses depend upon their proximity to the house. But the prices of neighboring houses are also observed only infrequently.

Let the log transaction price of dwelling  $i$  at time  $t$  be

$$V_{it} = P_t + Q_{it} + e_{it} = P_t + X_{it}\beta + e_{it}, \quad (1)$$

where  $V_{it}$  is the log of the observed transactions price of dwelling  $i$  at  $t$ , and  $P_t$  is the log of aggregate housing prices.  $Q_{it}$  is the log of housing quality, and can be parameterized by  $X_{it}$ , the set of housing attributes and by a set of coefficients,  $\beta$ , which price those attributes. If a transaction is observed at two points in time,  $t$  and  $\tau$ , and if the quality of the dwelling remains constant during the interval, then

$$\begin{aligned} V_{it} - V_{i\tau} &= P_t - P_\tau + (X_{it} - X_{i\tau})\beta + e_{it} - e_{i\tau} \\ &= P_t - P_\tau + e_{it} - e_{i\tau} \end{aligned} \quad (2)$$

With constant quality, (2) identifies price change in the market. Equation 2 also shows that the return on an individual dwelling can be decomposed into an aggregate return ( $P_t - P_\tau$ ) and an idiosyncratic return ( $e_{it} - e_{i\tau}$ ).

Let the idiosyncratic part of the house price (or the error term),  $e_{it}$ , consist of two components that are realized for each individual dwelling at the time of transaction:  $\eta_{it}$ , an idiosyncratic innovation without persistence; and  $\varepsilon_{it}$ , an idiosyncratic innovation with persistence,  $\varepsilon_{it} = \lambda\varepsilon_{it} + \mu_{it}$ . In addition, assume that the value of any particular dwelling depends also on innovations that occur to other dwellings

contemporaneously. We assume this spatial correlation depends on the distance between units.

$$e_{it} = \rho \sum_{j=1}^N w_{ij}e_{jt} + \xi_{it} = \rho \sum_{j=1}^N w_{ij}e_{jt} + \varepsilon_{it} + \eta_{it} = \rho \sum_{j=1}^N w_{ij}e_{jt} + \lambda\varepsilon_{i,t-1} + \eta_{it} + \mu_{it}, \tag{3}$$

where  $w_{ij}$  is some function of the distance between unit  $i$  and  $j$  and  $N$  is the number of dwellings in the economy. Let  $E(\eta_{it}\eta_{jt}) = 0, E(\varepsilon_{it}\varepsilon_{jt}) = 0, E(\eta_{it}^2) = \sigma_\eta^2$  and  $E(\mu_{it}^2) = \sigma_\mu^2$ . The value of a particular dwelling depends, not only on its own past and contemporaneous innovations, but also on innovations of other dwellings, past and contemporaneous.

In vector notation, expression (3) is

$$\mathbf{e}_t = \rho \mathbf{W} \mathbf{e}_t + \boldsymbol{\xi}_t, \tag{4}$$

where  $\mathbf{e}_t$  is a vector of  $e_{it}$  for all the dwellings at time  $t, \mathbf{e}_t = [e_{1t}, e_{2t}, \dots, e_{Nt}]'$ ,  $\mathbf{W}$  is a weight matrix, some measure of the distance between dwellings, and  $\boldsymbol{\xi}_t$  a vector of  $\xi_{it}$  where  $\boldsymbol{\xi}_t = [\xi_{1t}, \xi_{2t}, \dots, \xi_{Nt}]'$ , for all dwellings. By solving for  $\mathbf{e}_t$  and taking the difference between two transactions at times  $t$  and  $s$ , we have

$$\mathbf{e}_t - \mathbf{e}_s = (\mathbf{I} - \rho \mathbf{W})^{-1} (\boldsymbol{\xi}_t - \boldsymbol{\xi}_s). \tag{5}$$

The variance-covariance matrix of (5) is

$$E[(\mathbf{e}_t - \mathbf{e}_s)(\mathbf{e}_t - \mathbf{e}_s)'] = (\mathbf{I} - \rho \mathbf{W})^{-1} E[(\boldsymbol{\xi}_t - \boldsymbol{\xi}_s)(\boldsymbol{\xi}_t - \boldsymbol{\xi}_s)'] (\mathbf{I} - \rho \mathbf{W})^{-1}. \tag{6}$$

Equations 5 and 6 indicate that when the prices of dwellings are autocorrelated over time and space, the price of any unit in the market at any period will be predictably related to those of other units at other periods.

Transactions on dwellings occur only infrequently. First, consider the covariance in errors between a dwelling  $i$  sold at  $t$  and  $s$  and another dwelling  $k$  sold at  $\tau$  and  $\varsigma, E[(e_{it} - e_{is})(e_{k\tau} - e_{k\varsigma})]$ . From Equation 6, and letting  $\Pi = (\mathbf{I} - \rho \mathbf{W})^{-1} = [\pi_1, \pi_2, \dots, \pi_N]$ , we have

$$E[(e_t - e_s)(e_\tau - e_\varsigma)'] = \begin{bmatrix} \pi_1' \\ \pi_2' \\ \vdots \\ \pi_N' \end{bmatrix} \left\{ E[(\boldsymbol{\xi}_t - \boldsymbol{\xi}_s)(\boldsymbol{\xi}_\tau - \boldsymbol{\xi}_\varsigma)'] \right\} [\pi_1, \pi_2, \dots, \pi_N]. \tag{7}$$

The elements of this expression are,

$$E[(e_{it} - e_{is})(e_{k\tau} - e_{k\varsigma})] = \pi_i' \left\{ E[(\boldsymbol{\xi}_\tau - \boldsymbol{\xi}_s)(\boldsymbol{\xi}_\tau - \boldsymbol{\xi}_\varsigma)'] \right\} \pi_k. \tag{8}$$

Since  $E(\xi_{it}\xi_{js}) = 0$  for  $i \neq j, E[(\boldsymbol{\xi}_t - \boldsymbol{\xi}_s)(\boldsymbol{\xi}_\tau - \boldsymbol{\xi}_\varsigma)']$  is a diagonal matrix such that

$$E[(\boldsymbol{\xi}_t - \boldsymbol{\xi}_s)(\boldsymbol{\xi}_\tau - \boldsymbol{\xi}_\varsigma)'] = E[(\xi_{it} - \xi_{is})(\xi_{i\tau} - \xi_{i\varsigma})] \times \mathbf{I}. \tag{9}$$

It is straightforward to show that

$$E(\xi_{it}\xi_{it}) = \lambda^{|t-\tau|} \left( \frac{\sigma_\mu^2}{1-\lambda^2} \right) + 1(t=\tau)\sigma_\eta^2, \quad (10)$$

which implies

$$\begin{aligned} & E[(\xi_{it} - \xi_{is})(\xi_{it} - \xi_{is})] \\ &= (\lambda^{|t-\tau|} - \lambda^{|t-\varsigma|} - \lambda^{|s-\tau|} + \lambda^{|s-\varsigma|}) \left( \frac{\sigma_\mu^2}{1-\lambda^2} \right) \\ &+ [1(t=\tau) - 1(t=\varsigma) - 1(s=\tau) + 1(s=\varsigma)]\sigma_\eta^2. \end{aligned} \quad (11)$$

Therefore, the variance-covariance matrix of innovations between a dwelling  $i$  sold at  $t$  and  $s$  and another dwelling  $k$  sold at  $\tau$  and  $\varsigma$  is

$$\begin{aligned} E[(e_{it} - e_{is})(e_{k\tau} - e_{k\varsigma})] &= \pi'_i \{ E[(\xi_{it} - \xi_{is})(\xi_{it} - \xi_{is})] \times \mathbf{I} \} \pi_k \\ &= \left\{ (\lambda^{|t-\tau|} - \lambda^{|t-\varsigma|} - \lambda^{|s-\tau|} + \lambda^{|s-\varsigma|}) \left( \frac{\sigma_\mu^2}{1-\lambda^2} \right) \right. \\ &\left. + [1(t=\tau) - 1(t=\varsigma) - 1(s=\tau) + 1(s=\varsigma)]\sigma_\eta^2 \right\} \pi'_i \pi_k \end{aligned} \quad (12)$$

Equation 12 indicates how the variance-covariance matrix of residuals from the regression specified in (2) can be used to identify the temporal and spatial components of house price persistence,  $\lambda$  and  $\rho$ , respectively. Identification requires observing at least two transactions for each dwelling and observing the distance of each dwelling from all others in the market.

Note that this model of housing prices specializes to that of Bailey et al. (1963) when  $\lambda = \rho = 0$ , to that of Case and Schiller (1987) when  $\lambda=1, \rho=0$  and to those of Hill et al. (1999) and Englund et al. (1999) when  $\rho=0$ . When,  $\rho=0$  so that no spatial correlation is present, the variance of the return on an individual dwelling between  $t$  and  $s$  is

$$\text{var}(V_{it} - V_{is}) = \frac{2\sigma_\mu^2}{1-\lambda^2} (1 - \lambda^{t-s}) + 2\sigma_\eta^2, \quad (13)$$

which is concave in the transaction interval.

In Case and Shiller's (1989) model, with  $\lambda=1$  and  $\rho=0$ , the error term in individual housing price follows  $e_{it} = \varepsilon_{it} + \eta_{it}$  where  $\varepsilon_{it} = \varepsilon_{i,t-1} + \mu_{it}$ . Then, the variance of the return on an individual dwelling is

$$\text{var}(V_{it} - V_{is}) = (t-s)\sigma_\mu^2 + 2\sigma_\eta^2. \quad (14)$$

The variance increases linearly with the length of the time interval between transactions. Thus, with mean reversion in the data, a model based on a random walk assumption underestimates the return variances for housing transactions over short intervals, but overestimates the variances for housing transactions over long intervals.<sup>1</sup> The housing price indexes published by U.S. government agencies

<sup>1</sup> This is because the return variance is concave when the dwelling price follows a spatio-temporal correlation process (or a mean reverting process).

(e.g., the OFHEO, now FHFA, price indices for metropolitan areas) are based upon the repeat transactions model developed above with  $\lambda=1$  and  $\rho=0$ . But the computation procedures do include a second order term in the variance estimation (See Abraham and Schauman 1991; Calhoun 1996), so that the variance increases at a diminishing rate with the time interval between transactions ( $A < 0$ ).

$$\text{var}(V_{it} - V_{is}) = A(t - s)^2 + B(t - s) + C \quad (14')$$

## Data

The analysis below is based upon all private condominium transactions in Singapore during an eleven-year period. Non-landed properties (apartments and condominiums) account for roughly two-thirds of the Singapore housing stock, and units in condominiums account for almost 40% of private residential housing in land-scarce Singapore.<sup>2</sup>

The data include all transactions involving condominium dwellings during the period from January 1, 1990 to December 31, 2000.<sup>3</sup> An extensive set of physical characteristics of the dwellings is recorded. The date of the transaction is recorded as well as the date of completion of construction. In addition, the address, including the postal code, is reported. The postal code identifies the physical location—the block of flats or, quite often, the specific building. A matrix of distances among Singapore's fifteen hundred postal codes permits each dwelling to be located spatially. The data set includes transactions among dwellings in the standing stock, sales of newly constructed dwellings, and presales of dwellings under construction (where transactions may be consummated several months before the date construction is actually completed).

The panel nature of the data permits us to distinguish dwellings sold more than once, and this identifies the models specified in “A Micro Model of House Prices.” By confining the sample to dwellings in multifamily properties, we eliminate the types of dwellings for which additions and major renovations are feasible. The sample of multifamily dwellings is thus less likely to include those for which the assumption of constant quality between transactions (see Eq. 2) is seriously violated.

Singapore data offer another advantage in estimating the model of housing prices, namely a spatial homogeneity of local public services (e.g., police protection, neighborhood schools), especially when compared to cities of comparable size in North America. During the decade of the 1990s, there was no discernible trend in the quality of neighborhood attributes of the bundle of housing services.<sup>4</sup>

Table 1 presents a summary of the repeat sales data used in the empirical analysis reported below. There are several points worth noting. First, confirming the

<sup>2</sup> See Sing (2001) for an extensive discussion of the condominium market in Singapore.

<sup>3</sup> The data have been supplied by the Singapore Institute of Surveyors and Valuers (SISV) which gathers transactions data from a variety of sources including legal registration records and developers' transactions records.

<sup>4</sup> One possible exception to this may be accessibility, where improvement in the transport system and its pricing may have altered the workplace access of certain neighborhoods.



**Table 1** Summary of Transactions Data on Singapore Condominiums 1990–2000

Number of Times Sold	Number of Dwellings	Total Number of Transactions	Price		Average Interval			Distance to <sup>g</sup>		
			Average <sup>a</sup>	Std of Appreciation <sup>b</sup>	Average Appreciation <sup>c</sup>	Between Transactions <sup>d</sup>	Pretransaction Interval <sup>e</sup>	Average Size <sup>f</sup>	Number of Rooms	Nearest Subway Station
1	42,169	42,169	861			31.98	129.93	2.77	1.421	9,239
2	8,791	17,582	913	0.52%	0.71%	47.77	137.44	2.74	1.437	8,577
3	1,195	3,585	1,030	0.68%	1.13%	28.33	154.42	2.76	1.500	7,464
4	190	760	1,087	0.73%	1.29%	20.88	159.51	2.63	1.383	6,838
5	28	140	1,418	0.92%	1.53%	15.23	208.90	2.90	1.427	4,633
6	4	24	1,129	0.86%	1.60%	15.85	187.40	2.80	1.362	6,296

<sup>a</sup> Thousands of current Singapore Dollars

<sup>b</sup> Average price appreciation between transactions divided by average interval between transactions in months

<sup>c</sup> Standard deviation of price appreciation between transactions divided by average interval between transactions in months

<sup>d</sup> Number of months

<sup>e</sup> Average number of months from transactions to completion of construction of dwellings

<sup>f</sup> Average size of dwellings in square meters

<sup>g</sup> Average distance in kilometers

infrequency of housing transactions, the number of dwellings sold more than once is less than 20% of the population of dwellings sold during the eleven-year period. Only 3% of the 52,337 dwellings were sold more than twice in the eleven-year period.

Second, the average selling prices tend to be higher for dwellings sold more frequently. The rate of appreciation is also higher. On average, dwellings sold five times appreciate almost twice as fast as dwellings sold only twice. For the dwellings sold more frequently, price appreciation tends to be more volatile. Transactions involving high-turnover dwellings are apparently riskier, but this risk is compensated by higher returns.

Third, the intervals between transactions are longer for dwellings sold infrequently. In part, this is an artifact of the fixed sampling framework. For presold dwellings, the average elapsed time between transaction and completion of construction is largest for those sold least frequently, which is not consistent with the popular belief in Singapore that presales are associated with “speculation” in the housing market.

Fourth, there are some differences in the characteristics of the dwellings sold more frequently. They tend to be larger in area, contain more rooms, and they are more centrally located. Their transit access is similar to that of dwellings sold less frequently.

The data on condominium transactions supports a price index regression model of the form of Eq. 2,

$$V_{it} - V_{is} = P_t D_{it} - P_s D_{is} + \gamma \kappa_{it} - \gamma \kappa_{is} + e_{it} - e_{is}, \quad (15)$$

where  $D_{ij}$  is a variable with a value of 1 for the month  $j$  in which condominium  $i$  is sold and zero in other months, and  $P_j$  is the estimated coefficient for this variable. There are 132 of these time variables, one for each month between 1990 and 2000. If dwelling  $i$  has been presold,  $\kappa_{it}$  is the time interval between the transaction date and the completion of construction.<sup>5</sup> For dwellings sold after completion of construction,  $\kappa_{it}$  is zero. Thus, the estimated coefficient  $\gamma$  measures the monthly discount rate for presold dwellings, i.e., the discount for unrealized service flows from dwellings which have been purchased but which are not yet available for occupancy. The purchase of a dwelling before completion, or even before construction, is not unique to Singapore and has become rather common in condominium transactions, for example in vacation properties in the U.S. Pre-sale contracts provide liquidity to developers and insurance to consumers against unanticipated price increases in the market.<sup>6</sup> Of the 11,883 pairs of transactions noted in Table 1, 305 consist of presale pairs. For another 5,204 pairs, the first transaction was made sometime before the property was completed.

<sup>5</sup> Some dwelling units were presold multiple times before completion of construction. For those presold units, dates of contract are used for dates of transactions.

<sup>6</sup> Presales are widely employed in China, for example, to finance development of new housing estates. See Deng and Liu (2008)

## The Diffusion of House Price Innovations

We assume the error terms in Eq. 3,  $\eta_{it}$  and  $\mu_{it}$ , are normally distributed. The log likelihood function for the observed sample of condominium transactions is thus

$$\log L(P, \gamma, \lambda, \rho, \sigma_\eta^2, \sigma_\mu^2) = -\frac{T}{2} \log(2\pi) + \frac{1}{2} \left\{ \log(|\Omega|^{-1}) - (\delta' \Omega^{-1} \delta) \right\}, \quad (16)$$

where  $\delta = [V_{it} - V_{is} - P_t D_{it} + P_s D_{is} - \gamma \kappa_{it} + \gamma \kappa_{is}]$ ,  $\Omega = [\omega_{ij}]$ , and  $\omega_{ij} = E[(e_{it} - e_{is})(e_{jt} - e_{js})]$  defined in (15). We estimate the parameters,  $\lambda$ ,  $\rho$ ,  $\gamma$ ,  $P_t$ ,  $\sigma_\eta$  and  $\sigma_\mu$ , by maximizing the log likelihood (16), based on 11,883 observations of repeat sales of 10,288 dwellings sold two or more times. In (3), the weights are assumed to be inversely related to distance, up to 250 m.<sup>7</sup> The influence of any transaction extends for roughly 200,000 square meters in the surrounding area.

Table 2 reports the estimated error structure when it is assumed that the price of an individual dwelling follows a spatio-temporal correlation process, a mean reverting process, and a random walk process, respectively. In the most general model, Column A, the estimated serial correlation coefficient,  $\rho$ , implies a large persistence in individual housing prices, with a half life of more than 6 months. The estimated spatial correlation coefficient, 0.55, implies a slow spatial diffusion. These coefficients are quite precisely estimated; the estimated value for  $\rho$ , 0.89, is significantly different from one by a wide margin. The estimated coefficient for  $\gamma$ , the discount for the period between transaction and dwelling completion (for presold units), is 14 basis points. This represents a 1.7% annual discount for a dwelling unit sold today for occupancy a year hence. The magnitude of the discount is not trivial. During this period aggregate housing prices rose, on average, by 0.4% monthly; thus, the discount for presold units reduced the net price appreciation for consumers by one third.<sup>8</sup>

The second column reports parameter estimates for the model when individual house prices are allowed to follow a mean reverting process, but with  $\rho$  assumed to be zero. The estimated serial correlation coefficient,  $\lambda$ , is 0.72, somewhat smaller than the estimate in Column A. Likelihood ratio tests reject a random walk in house prices ( $\lambda=1$ ) and serially uncorrelated house prices ( $\lambda=0$ ) by wide margins,  $\chi^2=16,564$ , and  $\chi^2=1,670.6$  respectively. The estimated value of  $\lambda$  suggests that the half-life of a one-unit shock to housing prices is about 2 months.

The third column reports parameter estimates for a model in which individual house prices are assumed to follow a random walk process without spatial autocorrelation. Following OFHEO, the table reports the model with a quadratic term, as in (14'), which also fits the data better. These results suggest that the variance in house prices increases with transaction intervals up to 37 months.

Figure 1 presents estimated monthly price indexes derived from the three models reported in Table 2. The estimated price index from the model with a mean reverting

<sup>7</sup> This expedites computation of (16) by making the  $\Sigma$  matrix sparse. When  $\rho \neq 0$ , the  $\Sigma$  matrix is a full square matrix with a length equal to the number of observations. However, with some cutoff, the  $\mathbf{W}$  matrix is sparse. Since  $\pi_i$  is the  $i$ -th row of  $(\mathbf{I} - \rho \mathbf{W})^{-1}$ , when  $\mathbf{W}$  is sparse, most of the elements of  $\pi_i$  will be zero. This reduces computation time considerably. We experimented with several assumed cutoff values. They have no effect upon the results.

<sup>8</sup> It is possible that there might exist a discrete quality change when a dwelling unit is first used. In this case, the estimated  $\gamma$  would capture some of these effects as well.

**Table 2** Estimated Error Structure and Price Dynamics, Singapore condominium 1990–2000. (t-statistics in parentheses)

	A. General Spatial-Temporal Process: $\lambda \neq 1$ and $\rho \neq 0$	B. Mean Reverting Process: $\lambda \neq 1$ and $\rho = 0$	C. Random walk Process: $\lambda = 1$ and $\rho = 0$
$\lambda$	0.8877 (287.84)	0.7170 (62.022)	
$\rho$	0.5512 (713.29)		
$\sigma_\eta$	0.0535 (14.340)	2.82E-05 (0.0004)	0.1543 (6.0016*)
$\sigma_\mu$	0.0691 (46.606)	0.1214 (53.760)	0.0307 (2.4826*)
$\gamma$	-0.0014 (7.4748)	-0.0009 (6.2087)	-0.0015 (10.695)
Squared Interval			-1.26E-5 (3.1868)
Average of changes in log of price index	0.00384	0.00437	0.00387
Std. Dev. of changes in log of price index	0.03530	0.04471	0.05128
Loglikelihood	1112.31	277.00	-142.32
Mean Square Error	0.05770	0.05708	0.05727

Estimates are based upon maximizing log likelihood function in Eq. 16. The model also includes 132 time variables, one for each month between 1990 and 2000

\* T-statistics are reported for the variances, not the standard deviations

process and the index from the model with a spatio-temporal correlation process appear to move quite closely. The estimated price index from the random walk model is consistently lower than that implied by the other two models.

### The Course of Housing Prices, Investment Returns and Their Predictability

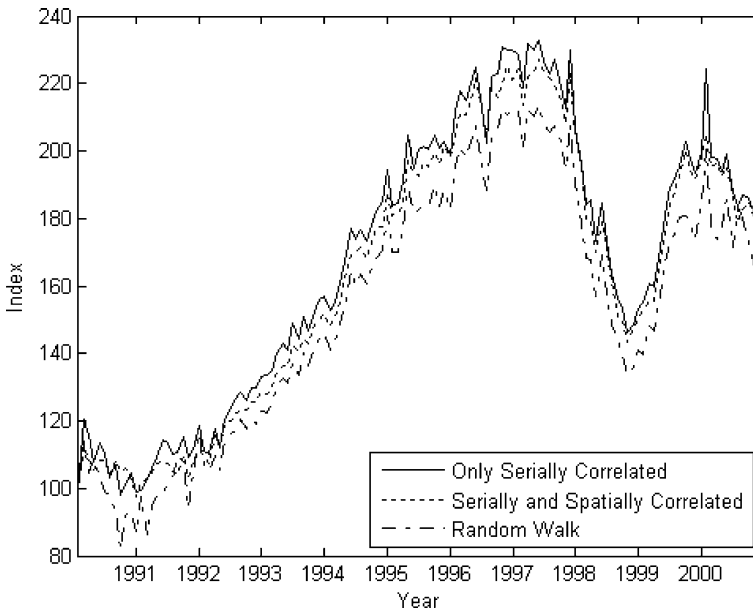
#### Predictability of Housing Returns

Although Fig. 1 reports similar patterns for the course of housing prices of Singapore dwellings, the investment returns implied by these aggregate indices are quite different. Ignoring transactions costs and leverage, the return in any period,  $R_t$ , is the change in the asset value plus the dividend (i.e., the rental stream,  $S_t$ , enjoyed during the period).

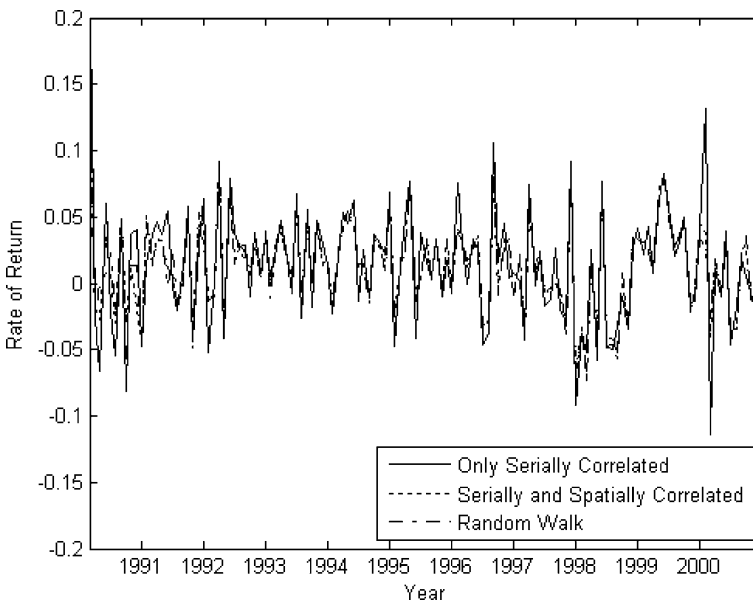
$$R_t = \left( \frac{P_t + S_t}{P_{t-1}} \right) \left( \frac{I_{t-1}}{I_t} \right), \tag{17}$$

where  $I_t$  is an index of the cost of living, less housing.

Figure 2 uses the non-housing component of the CPI for Singapore to chart the course of real investment returns in logarithms during the eleven-year period, from



**Fig. 1** Singapore Condominium Price Indices 1990–2000 (The figure graphs the monthly index values  $I_t = \exp[\sum_{j=1}^t P_j]$  where the values of  $P$  are estimated from models reported in Table 2 by maximizing the likelihood function in Eq. (16))



**Fig. 2** Monthly Returns to Investment in Singapore Condominiums 1990–2000

the three models.<sup>9</sup> Although the mean returns differ by less than five basis points per month, the patterns of estimated returns and the estimated volatility from the three models are strikingly different.

Table 3 reports tests of the predictability of estimated monthly returns for the three indexes. We investigate the forecastability of returns based upon one-month and three-month lags. There is a consistent disparity in the predictability of investment returns implied by models based upon the three price generating processes. When spatial and serial correlations are recognized in the estimation of housing prices (Column A), there is no evidence that aggregate returns to housing investment are predictable. However, when housing returns are estimated from a mean reverting process without spatial correlation (Column B), standard tests reject the null hypothesis of no predictability in returns. Finally, when returns are estimated from the conventional random walk model, they are strongly predictable. The p-values for both tests are less than 0.5%. These results are consistent for both a one-period and a three-period distributed lag.<sup>10</sup> Indeed, the nonparametric kernel-based test (Hong 1996), reported in the last row, shows that the result does not rely upon any parametric specification of lag structure.

The most striking feature of the table is that the predictability of the aggregate housing returns gradually disappears as restrictive assumptions on the individual housing price generating process are relaxed. When the assumption of the random walk with no spatial diffusion is maintained, p-values of the test statistics are very small. When the random walk assumption is relaxed (but spatial diffusion is not allowed), p-values are larger; the null hypothesis of no predictability in aggregate housing returns is still rejected at least for some of the tests. When returns on individual dwellings are allowed to be dependent over time and over space, p-values of the test statistics are all large enough that the null hypothesis of no predictability in aggregate housing returns is not rejected in any of the tests. This implies that the well known predictability in housing returns may arise simply because *the underlying price index is inaccurately estimated* due to restrictive assumptions about the price generation process.

Why does the aggregate return, when estimated with the random walk and no spatial diffusion assumption, appear to be significantly predictable while it exhibits no predictability when it is estimated without such restrictions?

To understand this, let  $\widehat{R}_t = R_t^* + \zeta_t$ , where  $\widehat{R}_t$  is a regression-based estimate of the aggregate return,  $R_t^*$  is the true (unobserved) return, and  $\zeta_t$  is the estimation error. Consider a regression of the estimated aggregate return on its lagged term,  $\widehat{R}_{t+1} = \beta_0 + \beta_1 \widehat{R}_t + v_t$ . The value of the AR(1) coefficient,  $\beta_1$ , is

$$\beta_1 = \frac{\text{cov}(R_t^*, R_{t+1}^*) + \text{cov}(\zeta_t, \zeta_{t+1})}{\text{var}(R_t^*) + \text{var}(\zeta_t)} \quad (18)$$

For the autocorrelation coefficient,  $\beta_1$ , to be zero, it is required that: the true aggregate return be unpredictable,  $\text{cov}(R_t^*, R_{t+1}^*) = 0$ ; and the estimation error be not persistent,  $\text{cov}(\zeta_t, \zeta_{t+1}) = 0$ . Conversely, a non-zero estimate of the AR(1) coefficient,  $\beta_1$ , need not imply predictability in returns at all; it can arise from

<sup>9</sup> We assume the implicit rent on owner-occupied condominiums,  $S_t = 0.01P_t$  (See Englund et al. 2002).

<sup>10</sup> These results are also apparent in more aggregated, quarterly price index models, not presented here.

**Table 3** Forecastability<sub>n</sub> of Investment Returns, Singapore Condominiums, 1990–2000 (p-values in parentheses)  $\hat{R}_t = \beta_o + \sum_i \beta_i \hat{R}_{t-i} + v_t$

	A. General Spatial-Temporal Process: $\lambda \neq 1$ and $\rho \neq 0$	B. Mean Reverting Process: $\lambda \neq 1$ and $\rho = 0$	C. Random walk Process: $\lambda = 1$ and $\rho = 0$
<b>1 Month Lag:</b>			
$\beta_1$	-0.0872	-0.1718	-0.2516
F-test for $\beta_1 = 0$	1.0395 (0.3099)	4.4974 (0.0359)	9.0974 (0.0031)
Box-Ljung Test	0.9706 (0.3245)	3.9582 (0.0466)	8.4758 (0.0036)
<b>3 Month Lag:</b>			
$\sum_{k=1}^3 \beta_k$	0.2324	0.0326	-0.1765
F-test for $\beta_1 = \beta_2 = \beta_3 = 0$	2.3508 (0.0756)	2.6531 (0.0516)	3.5673 (0.0161)
Box-Ljung Test	4.6916 (0.1958)	7.0204 (0.0713)	12.0101 (0.0073)
<b>Nonparametric Test:</b>			
Kernel-based test by Hong (1996)	0.8589 (0.3904)	1.5906 (0.1117)	4.2699 (0.0000)

persistent forecast errors. This implies that the well-known findings that housing prices and housing returns are predictable may arise, simply by construction, if  $\text{cov}(\zeta_t, \zeta_{t+1})$  is not equal to zero.

The properties of  $\zeta_t$  determine estimated autocorrelation coefficient,  $\beta_1$ . Suppose we have a sample of  $M$  houses transacted in each of  $T$  periods. If  $\text{cov}(R_t^*, R_{t+1}^*) = 0$ , it can be shown that

$$E(\hat{\beta}_1) = \frac{E(\zeta_t \zeta_{t+1}) + 0(\frac{1}{T})}{\sigma_r^2 + \sigma_\varepsilon^2 + 0(\frac{1}{T})} + \frac{\{E(\zeta_t \zeta_{t+1}) + 0(\frac{1}{T})\} 0(\frac{1}{T})}{\{\sigma_r^2 + \sigma_\varepsilon^2 + 0(\frac{1}{T})\}^3} - \frac{\sigma_\varepsilon^2 E(\zeta_t \zeta_{t+1}) + 0(\frac{1}{T})}{\{\sigma_r^2 + \sigma_\varepsilon^2 + 0(\frac{1}{T})\}^2}, \tag{19}$$

where  $E(\zeta_t \zeta_{t+1}) \rightarrow 0$  for large  $M$ .

Equation 19 shows that  $E(\hat{\beta}_1)$  will converge to zero only if  $T$  is large enough and  $M$  is also large enough. The convergence of  $E(\zeta_t \zeta_{t+1})$  depends on the spatial structure of the housing market, and convergence is slow when prices of individual dwelling units are more correlated.<sup>11</sup>

We can evaluate the impacts of the sample sizes of  $T$  and  $M$  by simulation. Using Singapore’s spatial structure, the previously estimated parameters, and assuming

<sup>11</sup> For example, when returns on individual dwelling units are not spatially correlated, its convergence speed is  $1/M$ . But when their prices are correlated, and their correlation is inversely related with distance, the convergence speed is  $\log(M)/M$ .

unpredictable aggregate returns,  $\beta_1=0$ , we simulate the Singapore private condominium market in two different dimensions,  $M$  and  $T$ . First, we simulate the price of each house in the sample each month for 10 years. We then randomly select some fixed number of houses each month, reflecting an underlying liquidity level or “transactions frequency.” We use the transactions so sampled to compute a repeat sales housing price index (Eq. 2), incorrectly assuming a random walk and no spatial diffusion. These estimates are used to compute the index-based returns, the first order autocorrelation coefficient of housing returns,  $\hat{\beta}_1$ , and its t-ratio. We construct the base case of 1% transactions probability, i.e., for each month 1% of total dwellings are randomly selected for trading. Then we extend the base case, first with a higher transactions frequency of 5%, and second with a longer time series of 50 years and a one-percent transactions frequency. Note that these two cases will generate the same number of observations.

Figure 3 summarizes these simulations, replicated 100 times. The figure reports the distributions of the estimated coefficient ( $\hat{\beta}_1$ ) and the corresponding t-statistics. It reports three distributions, the first with one-percent transactions frequency and monthly observations for 10 years, the second with the one-percent transactions frequency but monthly observations for 50 years, and the third with five-percent transactions frequency and monthly observations for 10 years.

A transactions frequency of 1% per month is much higher than the turnover rate observed in virtually all housing markets.<sup>12</sup> A transactions frequency of 5% per month is close to the turnover rate observed in the U.S. stock market.<sup>13</sup> Thus, the 1% figure represents an illiquid market, and a 5% figure represents a very liquid market.

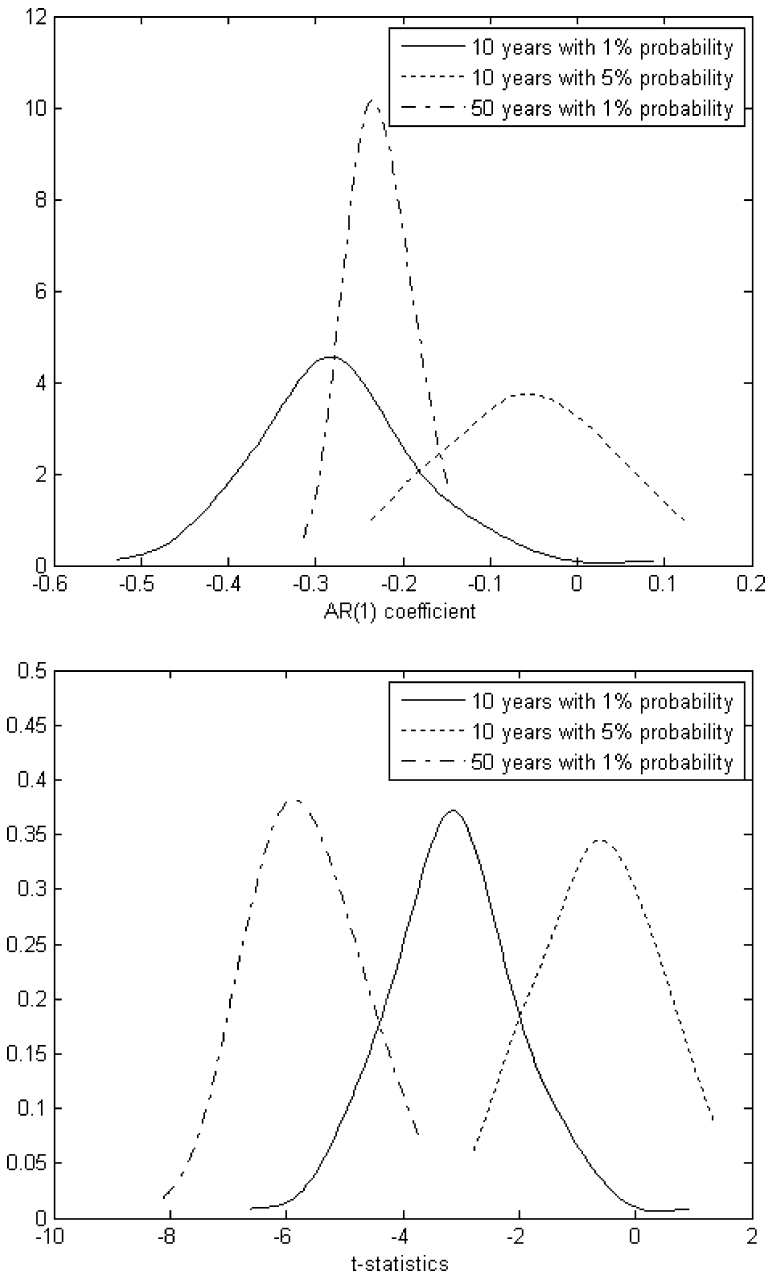
The probability distribution of  $\hat{\beta}_1$  with 1% transactions frequency and 120 monthly observations is sharply skewed to the left, centered around  $-0.3$ , quite far from the true value of 0. The distribution of t-statistics shows that usual t-tests are highly misleading. Among the 100 simulations, there are only fifteen instances where the t-statistic is larger than  $-2$ . This indicates that when the housing market has low transactions frequencies, even though aggregate housing returns are not predictable, it is very likely that the estimated returns will be highly predictable. This apparent predictability persists for low values of the monthly transactions frequency, and it is not eliminated until the monthly turnover rate reaches 5% level. When the transactions probability is 5%, the distribution is substantially further to the right, and the center of the distribution is quite close to zero; the t-statistics are between  $-2$  and  $2$  for 90 instances out of 100, and the null hypothesis of no predictability is rarely rejected using conventional tests.

However, when the sample period extends to 50 years (five times the original 10 years) with the transactions frequency kept at 1%, the results are quite different. The distribution moves to the right only by a small margin, and the center of the distribution is still well below  $-0.2$ . At the same time, dispersion of the distribution is reduced. In this instance, a small change in mean and a large reduction in variance

<sup>12</sup> The turnover rate in the Singapore market during the period was about 0.34 percent per month. The turnover rate for U.S. housing markets has averaged about 0.25 to 0.33 percent per month during the recent past. See Duca 2005.

<sup>13</sup> The weekly turnover rate of NYSE and AMEX during 1997–2001, reported in Cremers and Mei (2007), was 1.4%.





**Fig. 3** Empirical Distribution of AR(1) Coefficient and t-statistics With different transactions probabilities, in Singapore private condominium market

shifts the distribution of t-statistics further to the left. Therefore, if the transactions probability remains low, but the sample period is extended, the analyst is considerably more likely to reject the null hypothesis of no predictability. In fact, there is no instance out of 100 simulations where the t-statistics is larger than  $-2$ .

This simulation indicates how important it is to have enough observations cross sectionally ( $M$ ) in testing for the predictability of aggregate returns. At the same time, it also shows how difficult it is to test for the predictability of returns since the low frequency of housing transactions is not due to inadequate data collection, but arises as an inherent feature of housing markets.

## Investment Performance

The previous results show that persistent shocks in returns to individual dwellings over time and in space, coupled with infrequent housing transactions, leads to substantial small sample bias in the predictability of aggregate housing returns when measured under conventional repeat sales models. Is this economically important? How seriously do these biased forecasts affect investment decisions?

First, it is clear that the results noted above have separate implications for aggregate housing returns and for individual housing returns. Table 3 indicates that the aggregate housing return ( $R_t = P_t - P_{t-1}$ ) is not predictable at the level of the aggregate housing index, but the idiosyncratic housing return ( $e_{it} - e_{it-1} = [V_{it} - V_{it-1}] - [P_t - P_{t-1}]$ ) is still predictable. Let  $\phi$  denote the serial correlation of monthly returns in an individual dwelling. It is straightforward to show, from (12), that

$$\phi = -\frac{(1 - \lambda)\sigma_\mu^2 + (1 + \lambda)\sigma_\eta^2}{2\sigma_\mu^2 + 2(1 + \lambda)\sigma_\eta^2}. \quad (21)$$

For the Singapore housing market, the results in Table 3 indicate that an individual housing return is substantially persistent, and its monthly serial correlation is  $-0.29$ . This has significant implications for investment in the local housing market. A better knowledge of the price process for individual dwellings can lead to superior investment decisions in two ways. First, improvement may arise through better estimates of aggregate housing price trends. Different assumptions about the price generating process have small effects on the large-sample properties of slope coefficients, but, as shown above, they do have substantial effects on the efficiency of the estimated aggregate returns when transaction frequencies are low. An investor who relies on random walk and no spatial diffusion would conclude that the aggregate housing return is predictable. She would base her investment decision on the “pseudo-predictability,” which will likely lead to sub par performance. Second, improved performance will arise from basing investment decisions on more complete information. For example, when housing prices are spatially and serially correlated, knowledge of past and present innovations in nearby dwellings provides information valuable for predicting the future course of prices for any house that the investor considers for investment. The investor who believes in a random walk and no spatial diffusion in individual housing returns would simply ignore those signals. The empirical issue is whether these signals are economically important.

To explore this, we conduct a second simulation of investor activity which utilizes the structure of a hypothetical “Grid City.” In this hypothetical city, dwellings are located on a square grid with 41 points on a side, and a house at each interior point is

separated by 50 m from its four nearest neighbors. Assume that the price of each dwelling follows the spatial and temporal correlation process, as reported in Table 2 Column A (i.e.,  $\rho=0.55$  and  $\lambda=0.89$ ). We simulate the price of each house in the Grid City each month for 10 years. In the simulation, the true aggregate housing return is not predictable,  $\beta_1=0$ . We then analyze the results of investment rules which depend upon forecasts of future housing returns. The investment rule applied here is quite simple. Given assumptions on the price process and the consequent parameter values governing the processes, an investor makes forecasts for housing returns using all the available transactions information. The investor is instructed to “buy” if the forecasted return is greater than some preset threshold. When the investor decides not to buy, she is assumed to invest in some alternative asset that generates a risk free return. The threshold may be interpreted as some known transactions costs in the housing market. We set the risk free rate equal to zero for these simulations.

Transactions costs vary with housing market characteristics, financial market characteristics and tax systems, so it is difficult to specify a precise level.<sup>14</sup> We use 0%, 6% and 12% as investment thresholds, comparable with a wide range of plausible transactions and opportunity costs.<sup>15</sup>

The simulation of housing returns is performed in a similar manner to the previous section. The investment holding period is set at 24 months, 48 months, and 96 months. We assume that the investor observes the market and collects transactions information during an initial observation period. We set the observation period at 24 months, 48 months, 72 months and 96 months. For simplicity, we concentrate on the dwelling unit located at the center of the grid. A price for each of 1,681 dwelling units is generated for every month of the observation period and the holding period. From the simulated prices, a sample of houses is selected each month with the preset transactions probability of 1%. Using the observations on the houses in the sample, together with her estimates of the parameters ( $\lambda$  and  $\rho$ ), the investor makes a price forecast for the next 24, 48, or 96 months. If the forecasted return exceeds the threshold, she will invest. The price at the end of the holding period is then used to evaluate the return on her investment. We consider two investors with differing information. The fully informed investor is armed with the full knowledge of housing price dynamics which follows the correlation process reported in Table 2, column A,  $\rho=0.55$  and  $\lambda=0.89$ . The uninformed investor forms her own forecasts in a similar manner. However, she assumes  $\rho=0$  and  $\lambda=1$ , and does not recognize the serial and spatial correlation of prices. It is assumed that the transactions cost is paid when the house is sold, so the net performance of the investment is the capital gain less transactions cost.

Table 4 summarizes the forecast performance of the two investors. The table reports the average percent difference between the true price at the end of the holding period and the forecast made by the two investors. Each investor uses the information available in the observation period to make a forecast of the price at the end of the holding period. The forecast is compared to the actual price, and the average (absolute) percent deviations are

<sup>14</sup> The *ex post* opportunity cost of housing investment in Singapore during the period 1990–2000 was in any case quite low. (Annual stock market returns averaged 0.1 percent; Treasury bill yields were about the same).

<sup>15</sup> For a more systematic examination of likely transactions costs in real estate, see Söderberg (1995) or Quigley (2002).

**Table 4** Difference in Forecasts by Fully Informed Investors and Uninformed Investors for Different Observation Periods and Holding Periods (2500 replications with transactions probabilities of 1% per month)

Entries in the table represent the average (absolute) percent difference between the true price at the end of the holding period and the forecast made by the informed and the uninformed investor

Observation Periods	Holding Period	Holding Period			
		6 month	24 months	48 months	96 months
24 months	Informed	15.85%	28.72%	46.68%	79.04%
	Uninformed	15.98%	28.54%	47.16%	82.48%
48 months	Informed	14.34%	22.26%	33.24%	51.92%
	Uninformed	15.68%	24.22%	34.16%	54.80%
72 months	Informed	13.91%	21.44%	28.92%	43.44%
	Uninformed	15.42%	23.88%	30.64%	46.00%
96 months	Informed	13.83%	20.96%	27.76%	40.80%
	Uninformed	14.96%	23.62%	30.12%	43.04%

reported in the table. The table reports the results of 2500 replications of this comparison using an underlying transactions probability of 1%.

Clearly the percentage errors are larger when the forecast is for prices further in the future (that is, when the assumed holding period is longer). The percentage errors are likewise smaller when the forecast is based upon more information (that is, when the forecasts are based upon a longer period of observing property transactions).

The results clearly establish that the informed investor makes better forecasts of future prices. For 15 out of 16 comparisons, the average error in the forecasts is less for the informed than for the uninformed investor. The difference is larger for shorter holding periods, but this advantage extends up to a holding period of 8 years.

The economic significance of the small, but systematic, advantage of the informed investor in forecasting is analyzed in Table 5. Table 5 reports the increased returns, in percentage points, to the informed investor as a function of the observation period and the holding period. The average increased return to the

**Table 5** Economic Advantage of Fully Informed Investor over Uninformed Investors in Percentage Points. (2500 replications with transactions probabilities of 1% per month)

Observation Periods	Holding Period			
	6 month	24 months	48 months	96 months
A. Transactions Cost: 0				
24 months	2.21%	1.96%	1.52%	0.96%
48 months	2.89%	2.64%	2.16%	1.68%
72 months	2.80%	3.20%	2.24%	1.20%
96 months	2.42%	3.58%	2.08%	1.44%
B. Transactions Cost: 6%				
24 months	1.77%	2.52%	2.12%	1.20%
48 months	1.68%	3.64%	2.28%	1.92%
72 months	1.92%	4.36%	2.64%	1.52%
96 months	1.57%	4.72%	3.32%	1.68%
C. Transactions Cost: 12%				
24 months	-0.24% <sup>a</sup>	2.38%	2.36%	1.04%
48 months	0.13% <sup>a</sup>	3.48%	2.60%	2.00%
72 months	0.33%	4.30%	3.24%	1.84%
96 months	0.20%	3.84%	3.60%	1.92%

<sup>a</sup>: Statistically insignificant at 5%

informed investor is reported for 2500 replications with an underlying transactions frequency of 1%. Results are reported for transactions costs of 0, 6, and 12%.

With no transactions costs, the informed investor earns a return that is about 200 basis points higher than the uninformed investor. Even with high transactions costs, the fully informed investor outperforms the uninformed investor by one to four percentage points. Only when transactions costs are very high (12%) and holding periods are very short (6 months), does the uninformed investor perform almost as well as the informed investor.

## Conclusion

For the past 15 years, it has been widely accepted that investment returns in housing are predictable. It is also widely believed that, due to high transactions costs, it is difficult to take advantage of this predictability. This paper develops a model of housing price determination that considers spatial correlation and serial correlation concurrently. Using comprehensive data on all Singapore condominium transactions, we estimate the extent of predictability in aggregate housing returns and in individual housing returns. The analysis supports a general model of price discovery, rejecting a simple random walk model as well as a model with the mean reversion without spatial correlation. More importantly, we find that assumptions about the underlying price process for individual dwellings affect the predictability in aggregate housing returns. There is no evidence of strong predictable aggregate housing returns when the appropriate price process—persistent idiosyncratic returns over time and space—is taken into account. This need not necessarily imply that housing market is efficient; individual housing returns are still persistent. The results indicate that the housing return is only predictable at the level of the individual asset, not at the aggregate level. In contrast, when the aggregate housing price index is estimated from a random walk model without spatial diffusion, the estimated aggregate housing return shows substantial predictability. We show that this arises from two sources; the illiquid nature of housing transactions and the persistent shocks on individual housing prices in time and over space. With these two factors present, the estimated predictability in aggregate housing returns contains substantial small sample bias. We show that the bias cannot be removed by extending time periods in the sample. More observations over time at the same transactions frequency only exacerbate this bias. The bias disappears only when the transactions frequency is high enough. However, for the bias to disappear, the transactions frequencies in the housing market would have to be as liquid as the transactions frequencies in stock markets.

Our simulation results suggest that an investor with enough information about the individual housing price process can, in fact, enjoy higher returns to housing investment. Our simulation results show that the investment performance of the fully informed investor is indeed superior to that of the naïve investor, even though her performance is bounded by holding periods and transactions costs.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Abraham, J. M., & Schauman, W. S. (1991). New evidence on home prices from Freddie Mac repeat transactions. *AREUEA Journal*, 19(3), 333–352.
- Bailey, M. J., Muth, R. F., & Nourse, H. O. (1963). A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58, 933–942.
- Calhoun, C. A. (1996). OFHEO house price indexes: HPI technical description, working paper, Office of Federal Housing Enterprise Oversight, Washington, D.C.
- Case, K. E., & Shiller, R. J. (1989). The efficiency of the market for single family houses. *American Economic Review*, 79(1), 125–137.
- Cremers, M., & Mei, J. (2007). Turning over turnover. *Review of Financial Studies*, 20, 1749–1782.
- Deng, Y., & Liu, P. (2008). Mortgage prepayment and default behavior with embedded forward contract risks in China's housing market. *Journal of Real Estate Finance and Economics*, forthcoming.
- Duca, J. (2005). Making sense of elevated housing prices. *Federal Reserve Bank of Dallas: Southwest Economy*, 5, 7–13.
- Englund, P., Gordon, T. M., & Quigley, J. M. (1999). The valuation of real capital: a random walk down kungsgaten. *Journal of Housing Economics*, 8(3), 205–216.
- Englund, P., Hwang M., & Quigley, J. M. (2002) Hedging Housing Risk. *Journal of Real Estate Finance and Economics*, 24(1), 163–197.
- Gatzlaff, D. H. (1994). Excess returns, inflation, and the efficiency of the housing markets. *Journal of the American Real Estate and Urban Economics Association*, 22(4), 553–581.
- Guntermann, K. L., & Norrbin, S. C. (1991). Empirical tests of real estate market efficiency. *Journal of Real Estate Finance and Economics*, 4(3), 297–313.
- Hill, R., Sirmans, C. F., & Knight, J. R. (1999). A random walk down main street. *Regional Science and Urban Studies*, 29(1), 89–103.
- Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica*, 64(4), 837–864.
- Hosios, A. J., & Pesando, J. E. (1991). Measuring prices in retransaction housing markets in Canada: evidence and implications. *Journal of Housing Economics*, 1(1), 1–15.
- Malpezzi, S. (1999). A simple error correction model of housing prices. *Journal of Housing Economics*, 8(1), 27–62.
- Quan, D. C., & Quigley, J. M. (1991). Price formation and the appraisal function in real estate markets. *Journal of Real Estate Finance and Economics*, 4(2), 127–146.
- Quigley, J. M. (2002). Transactions costs and housing markets. In T. O' Sullivan & K. Gibb (Eds.), *Housing Economics and Public Policy* (pp. 56–66). London and New York: Blackwell.
- Sing, T. F. (2001). Dynamics of the condominium market in Singapore. *International Real Estate Review*, 4(1), 135–158.
- Söderberg, B. (1995). Transaction costs in the market for residential real estate. *Department of Economics Working Paper #20*, Royal Institute of Technology, Stockholm.