

## INTERNATIONAL REAL ESTATE REVIEW

2010 Vol. 13 No.1: pp. 1 – 29

# **A Hierarchical Bayesian Approach for Residential Property Valuation: Application to Hong Kong Housing Market**

**Sam K. Hui\***

Stern School of Business of New York University; Email: [khui@stern.nyu.edu](mailto:khui@stern.nyu.edu)

**Alvin Cheung**

Massachusetts Institute of Technology

**Jimmy Pang**

Stanford Artificial Intelligence Laboratory, Stanford University

We have developed a statistical method for the valuation of residential properties using a hierarchical Bayesian approach, which takes into consideration the unique structure of the Hong Kong property market. Our model is calibrated on a dataset that covers all residential real estate transactions in ten major Hong Kong residential complexes from February 2008 to February 2009. Although parsimonious, our model outperforms other valuation methods that are based on average price-per-square-foot or expert assessments. By providing our model-based valuations online without charge, we hope to improve transparency in the Hong Kong housing market, thus enabling consumers to make better investment decisions.

---

\* Correspondence author

## 1. Introduction

Property valuation is an important issue for consumers, real estate investors, and the banking industry. For consumers and real estate investors, accurate assessments of residential real estate property values are crucial for financial planning and making investment decisions. For banks, there is a need to value residential properties accurately in order to make optimal lending decisions (e.g., evaluating the size of the mortgage as a percentage of the property value that the bank is willing to offer) and to assess the amount of risk in their existing mortgage portfolios (Lentz and Wang 1998).

This paper focuses on property valuation based on past transaction data in the Hong Kong residential property market, an active market for real estate transactions (Choy et al. 2007). The obtaining of accurate property valuation in the Hong Kong market based on previous transaction prices poses several key methodological challenges.

First, although considered an active market, data on Hong Kong real estate transactions are “sparse” in statistical terms. For instance, in our dataset (described in detail in Section 4) that covers ten major residential estates in Hong Kong (based on transaction volume), only around 5% of all property units record a transaction in the one-year period from February 2008 to February 2009. Thus, to obtain valuation for a property unit that does not have any transaction records, we need to utilize information from other transactions, i.e., obtain valuation information from “similar” units that had recorded transactions. To combat against the sparseness of real estate transaction data, a model-based valuation approach is necessary.

Second, the structure of the Hong Kong real estate market is different from their North American (e.g., Case and Shiller 1989) or European (e.g., Clapham et al. 2006) counterparts, where single-family homes are more common. As will be discussed in Section 2, most residential properties in Hong Kong are high-rise condominiums, with many units sharing the same floor plan within a building. Also, multiple condominium buildings are usually replicated from a common design to form a housing estate. In order to improve estimation efficiency, our model must explicitly account for this unique hierarchical structure.

Furthermore, the real estate market in Hong Kong is highly volatile and exhibits huge month-to-month price fluctuations. For instance, in our dataset that covers a one-year period from February 2008 to February 2009, the median per-square-foot price of a property changes by as much as 12.7% month over month. Thus, when modeling property values, one must account for not only the transaction history and the characteristics of the property unit, but also when those transactions occurred to adjust for general market conditions at that time.

In this paper, we address these methodological challenges by developing a hierarchical Bayesian model based on the framework proposed by Gelman et al. (2003) for property valuation in the Hong Kong housing market. Essentially, our proposed method is a Bayesian extension of the commonly used “hedonic pricing” models (i.e., multiple regression models) in the real estate literature (Tiebout 1956; Lancaster 1966; for a recent review, see Lentz and Wang 1998). Our Bayesian approach borrows information efficiently across different housing units through a hyperdistribution specification (Gelman et al. 2003), which guarantees reasonable estimation accuracy even when transaction records are sparse. Meanwhile, the hierarchical nature of our model takes into account the unique structure of the Hong Kong property market. In addition, we address the month-to-month fluctuations in general price levels using a flexible “random walk with drift” time series specification (Greene 2007). Through an out-of-sample assessment, we show that our model outperforms average price-per-square-foot- or expert assessment-based valuation methods.

The contributions of our paper are as follows. First, to the best of our knowledge, this is the first paper that takes into account, the unique features of the Hong Kong property market (e.g., multiple units sharing the same floor-plan and multiple replicas of the same building in an estate) using a hierarchical Bayesian model. In this respect, this paper is similar in spirit to other methodological papers, such as Bao and Wan (2004), and Zhou and Bao (2007), which developed and applied advanced statistical methods to model the Hong Kong residential real estate market. Second, given that our data collection period (February 2008 to February 2009) coincides with the peak of the global subprime crisis, our model parameters provide a complementary insight (in addition to conventional median-price trend analysis) on the effect of the global financial turmoil on Hong Kong real estate prices.

The remainder of this paper is organized as follows. Section 2 discusses the background information on the Hong Kong property market, and briefly reviews previous literature that models Hong Kong real estate prices. Section 3 develops our hierarchical Bayesian approach for home price valuation. Section 4 describes our dataset on real estate transactions. In Section 5, we apply our model on actual data, describe our model fit, and compare the out-of-sample valuation performance against other benchmark methods. Finally, Section 6 concludes with directions for future research.

## **2. Background: Hong Kong Housing Market**

The Hong Kong residential private property market is one of the most actively traded real estate markets in the world (Bao and Wan 2004). Residential real estate in Hong Kong is divided into public housing and private housing; roughly half of the seven-million population in Hong Kong live in public

housing, and the other half in private housing (Hong Kong Housing Authority 2008). In terms of housing supply, in 2008, there were a total of 2.5 million residential housing units in Hong Kong, among which 1.1 million are public housing units and 1.4 million are private property units (Hong Kong Housing Authority 2008). Similar to the focus of previous research on the Hong Kong housing market (e.g., Bao and Wan 2004; Chau et al. 2004; Choy et al. 2007; Mok et al. 1995), this paper focuses on valuating private property units.

Most of the private housing units in Hong Kong are high-rise condominiums organized into residential estate complexes. Figure 1 shows the general structure of housing units in Hong Kong. The left panel shows an estate complex (South Horizon), which is comprised of multiple (in this case, 34) “blocks,” as shown in the right panel. Each block is further divided into floors, and each floor is divided into multiple (in this case, 8) units. Note that different floors of the same unit usually have the same floor plan. This unique feature allows us to model the value of a unit by its estate, block, unit, and floor specifications.

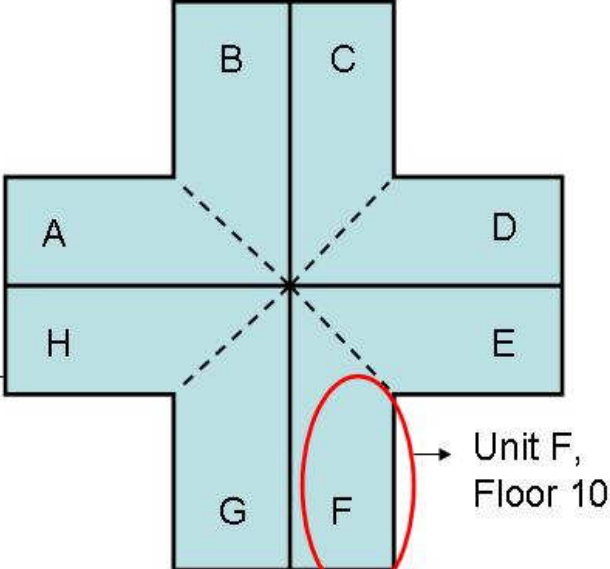
Previous literature on modeling the Hong Kong real estate market generally focuses on two distinct issues. One stream of literature attempts to create an overall property price index similar to the S&P/Case-Shiller real estate index (Case and Shiller 1989). Based on the methodology proposed by Bailey et al. (1963), Chau (2006) develops an index construction method based on same-unit repeated sales data. Similarly, Chan et al. (2009) create the Centa-City index to capture the general month-to-month overall price changes in the residential real estate market, while Hui and Yue (2006) analyze the time series of Hong Kong housing prices to investigate evidence of an asset “bubble”.

Another stream of research applies hedonic price models (e.g., Gillard 1981, Li and Brown 1980, Rosen 1974) to study the extent to which characteristics of housing units (e.g., square footage, age, and neighborhood) drive property transaction prices. For instance, Choy et al. (2007) apply a hedonic price model to transaction data “during a period of relatively stable property price” to study the effects of floor level and property size on property value, and find that larger units and higher floor levels generally command higher values. A similar hedonic-modeling approach is taken by Mok et al. (1995) to study the relationship between property prices and dwelling characteristics, while Bao and Wan (2004) use a smoothing spline technique to estimate hedonic price models and show that their estimation technique outperforms traditional least-square estimation methods.

**Figure 1** Example of a Housing Unit in Hong Kong



Estate: South Horizon



Block: 28

Unit F,  
Floor 10

Our Bayesian model integrates both streams of literature to address the issue of property valuations for all housing units, regardless of whether a transaction exists for a particular unit. While allowing for the estimation of a general price level for each estate (an index), our Bayesian approach also allows us to bring in dwelling characteristics to integrate information from one housing unit to another.

### 3. Hierarchical Bayesian Approach for Residential Property Valuation

This section develops our hierarchical Bayesian approach for residential property valuation in Hong Kong. Similar in spirit to the spatio-temporal modeling approach in Sun et al. (2005), we model property value by considering the effects of time, building-block, and floor. Section 3.1 discusses the details of our model. Section 3.2 outlines the prior specification and our computation procedure. Section 3.3 demonstrates the application of our model to property valuation.

#### 3.1 Model of Transaction Prices

Throughout this paper, we use the following set of notations. Let  $i$  ( $i = 1, 2, \dots, I$ ) denote estate complexes (e.g., South Horizon);  $j$  ( $j = 1, 2, \dots, m_i$ ) denote a block-unit within an estate (e.g., Block 28-Unit F);  $k$  ( $k = 1, 2, \dots, n_j$ ) denote the floor number (10th floor); and  $n_j$ , the number of floors in block-unit  $j$  minus one.<sup>1</sup>

Let  $y_{ijkt}$  be the log-transaction price of estate  $i$ , block-unit  $j$ , floor  $k$  at time  $t$  (measured in months).<sup>2</sup> Building upon the previous literature on hedonic pricing models (e.g., Choy et al. 2007), we model  $y_{ijkt}$  as the sum of three distinct model components; namely: (i) time effect, (ii) block-unit effect, (iii) floor effect, and a random error term. Formally,

$$y_{ijkt} = \underbrace{\alpha_{it}}_{\text{time effect}} + \underbrace{\phi_{ij}}_{\text{block-unit effect}} + \underbrace{\psi_{ijk}}_{\text{floor effect}} + \underbrace{\varepsilon_{ijkt}}_{\text{random error}}, \quad \varepsilon_{ijkt} \sim N(0, \sigma_{(i)y}^2) \quad (1)$$

<sup>1</sup> Based on discussions with real estate agents, the top-floor units can be systematically different from those on other floors. For instance, some top-floor units include exclusive access to the roof-terrace area, and some tenants install (sometimes illegal) “upgrades” to the roof terrace, thereby affecting their values. Thus, in this paper, we exclude the top floor units from our empirical analysis.

<sup>2</sup> Note that besides the logarithmic transformation, other types of transformations (e.g., Box-Cox transformation) can also be considered (see Cropper et al. 1988).

We now explain the three model components in detail. First,  $\alpha_{it}$  represents the time effect for estate  $i$  at time  $t$ . This captures any general changes in market values for all block-units within estate  $i$  at time  $t$  due to exogenous economic/demographic changes in the neighborhood (e.g., changes in macro-economic conditions, such as real interest rates, changes in transportation networks, opening of new stores and schools in the area). Building on the previous literature (Case and Shiller 1989), we model  $\alpha_{it}$  using a random-walk-with-drift time series specification<sup>3</sup> (Greene 2007), with drift parameter  $\mu_i$  and variance parameter  $\sigma_{(i)\alpha}^2$ :

$$\alpha_{it} = \alpha_{i(t-1)} + w_{it}, \quad w_{it} \sim N(\mu_i, \sigma_{(i)\alpha}^2) \quad (t > 1) \quad (2)$$

This specification is widely used in econometric models of non-stationary time series (e.g., Greene 2007), such as gross national product (Cochrane 1988), personal income (Mankiw and Shapiro 1985), and foreign exchange rates (Engel and Hamilton 1990). It is flexible and assumption-free, yet maintains reasonable smoothness in the time series; it also provides a conjugate prior to our model estimation procedure using Markov chain Monte Carlo (MCMC) sampling (Choi et al. 2009), as described in Appendix I. In relation to the previous literature on price indexes,  $\alpha_{it}$  can be interpreted as the “price index” for estate  $i$  at time  $t$  (Messe and Wallace 1997), which represents the general market conditions over time, controlling for the characteristics of housing units. We will revisit this interpretation when we discuss our empirical results in Section 5.2.

Second,  $\phi_{ij}$  denotes the block-unit effect for the  $j$ -th block-unit in estate  $i$ . Following hedonic price models (e.g., Ball 1973; Bao and Wan 2004; Chau et al. 2001; Choy et al. 2007; Freeman 1979; Leggett and Bockstael 2000), we allow  $\phi_{ij}$  to be driven by the intrinsic features of the property unit (e.g., square footage, view, age), using a hierarchical specification as follows:

$$\phi_{ij} = \bar{x}_{ij}' \beta_i + \delta_{ij}, \quad \delta_{ij} \sim N(0, \sigma_{(i)\phi}^2) \quad (3)$$

where  $\bar{x}_{ij}$  denotes the vector of features for block-unit  $j$  in estate  $i$ ;  $\beta_i$  is a vector of parameters that measures the sensitivity of property values on these features.

---

<sup>3</sup> We check the validity of the random-walk-with-drift assumption using autocorrelation plots (of different lags) of the  $w_{it}$  time series, which are included in Appendix II; we find that none of the autocorrelations are significant, which provides some empirical support for our model specification.

We should note that our model specification is considerably more flexible than commonly used hedonic price models. Instead of assuming that the price of a housing unit is solely driven by a set of features ( $\bar{x}_{ij}$ ), Equation (3) specifies a prior distribution on the block-unit effect  $\phi_{ij}$  conditioned on these features. This allows our model to effectively “borrow information” across different block-units (while controlling for their different characteristics) using a hierarchical modeling framework (Gelman et al. 2003).

Third,  $\psi_{ijk}$  denotes the floor effect. Previous literature on Hong Kong property prices (e.g., Bao and Wan 2004; Choy et al. 2007) typically find that higher floor units command higher market values. Thus, we model the floor effect as a linear function of floor level, as follows:

$$\psi_{ijk} = \gamma_i k \quad (4)$$

One may argue that the linearity assumption in Equation (4) is rather restrictive. It is possible that floor effect can be non-linear, as suggested by Choy et al. (2007). However, as will be discussed in Sections 5.1 and 5.3, we find that our parsimonious specification already provides an excellent description of the data both in- and out-of-sample. In Section 6, we generalize Equation (4) by relaxing the linearity assumption.

Finally, the random error terms  $\varepsilon_{ijkt}$  denote any unobserved effects beyond our model specification. This includes, for instance, the bargaining power of the buyer and seller, and any unobserved conditions of the housing unit, e.g., renovations and upgrades (Bailey et al. 1963).

### 3.2 Prior Specification and Computation Procedure

To complete our hierarchical Bayesian specification, we specify conjugate, weakly informative priors for all model parameters to allow for efficient posterior sampling using the Gibbs sampler (Casella and George 1992), and at the same time, allow the data to dominate the posterior inference. Specifically, we apply the following sets of standard, diffuse prior distributions for our model parameters (Gelman et al. 2003):

$$\alpha_{i1} \sim N(0, 100^2) \quad (5)$$

$$\beta_i \sim N(0, 100^2 I) \quad (6)$$

$$\gamma_i \sim N(0, 100^2) \quad (7)$$

$$\mu_i | \sigma_{(i)\alpha}^2 \sim N(0, 100^2 \sigma_{(i)\alpha}^2) \quad (8)$$

$$\sigma_{(i)\alpha}^2, \sigma_{(i)\phi}^2, \sigma_{(i)y}^2 \sim Inv - \chi^2(0.001, 1) \quad (9)$$

where  $i$  represents the identify matrix.



Given the prior distributions in Equations (5) – (9), our model specification is complete. We then obtain the posterior distributions of our parameters using the Gibbs sampler. Since the Gibbs sampling procedure described here is standard in the Bayesian statistics literature, we briefly outline each step in Appendix I. Readers may refer to Casella and George (1992) for details.

### 3.3 Model-Based Property Valuation

Once we obtain the posterior distribution for our model parameters, it is straightforward to generate the valuation for each property unit, regardless of whether or not the unit has had any previous transactions. More specifically, let the (log-) market value of unit  $j$ , on floor  $k$ , in estate  $i$  at time  $t$  be  $\theta_{ijkt}$ .

Our model-based estimate for  $\theta_{ijkt}$  (denoted as  $\hat{\theta}_{ijkt}$ ) is the posterior mean<sup>4</sup> given all the data:

$$\hat{\theta}_{ijkt} = E(\theta_{ijkt} | Y) = E(\alpha_{it} + \phi_{ij} + \gamma_i k | Y) \quad (10)$$

where  $Y$  denotes all historical transactions across all units. Thus, the model-based valuation (and the associated posterior interval) for any property unit can be easily estimated by sampling from the posterior distributions of  $\alpha_{it}$ ,  $\phi_{ij}$ , and  $\gamma_i$  using MCMC methods described in Appendix I. Formally,

$$\hat{\theta}_{ijkt} \approx \frac{1}{M} \sum_{m=1}^M (\alpha_{it}^{(m)} + \phi_{ij}^{(m)} + \gamma_i^{(m)} k) \quad (11)$$

where  $\alpha_{it}^{(m)}$ ,  $\phi_{ij}^{(m)}$ ,  $\gamma_i^{(m)}$  denotes the  $m$ -th sample from the posterior distribution of  $\alpha_{it}$ ,  $\phi_{ij}$ , and  $\gamma_i$ , respectively. The posterior interval for  $\theta_{ijkt}$  can likewise be computed from the posterior distributions of the model parameters.

## 4. Data Overview

We obtained our dataset from a major real estate agent in Hong Kong. It contains all transactions in ten major residential estate complexes (based on the highest transaction volume during the period) in Hong Kong between February 2008 and February 2009 as recorded by the Hong Kong Government Land Registry. We have focused on the above time period because it coincides with the global financial crisis, allowing us to study the impact of the financial crisis on the Hong Kong housing market. We will return to this issue when we discuss our results in Section 5.2. We have also obtained the

---

<sup>4</sup> The posterior median can also be considered. Given that posterior distributions in our application is fairly symmetric, both estimates provide similar results. Details are available upon request.

size (in square feet) of each unit within the estates. The list of the ten major estate complexes, along with a summary of their features, is shown in Table 1. The largest estate (in terms of the number of housing units) in our sample is Kingswood Villas, with nearly 16,000 housing units; the smallest estate is Amoy Garden, with just over 5,000 units. The interior size of the housing units ranges from 373 square feet to 1920 square feet.

**Table 1 List of Ten Major Estates and Their Characteristics**

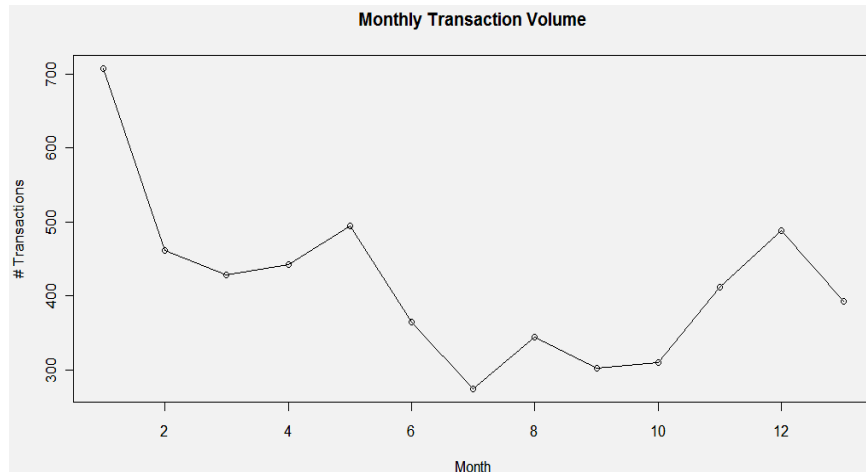
<b>Estate name</b>	<b>Number of transactions</b>	<b>Number of housing units</b>	<b>Smallest size (sqft)</b>	<b>Largest size (sqft)</b>
<b>Taikoo</b>	469	13576	585	1237
<b>South Horizon</b>	413	9856	632	1121
<b>Kornhill</b>	266	6584	582	1056
<b>Mei Foo</b>	698	14385	560	1920
<b>Whampoa</b>	363	11120	469	1110
<b>Laguna City</b>	367	8080	639	941
<b>Amoy Garden</b>	394	5128	371	607
<b>Kingswood</b>	1170	15888	573	825
<b>City One</b>	822	10970	389	1018
<b>Metrocity</b>	460	6768	487	1026

The ten major estates contain a total of 5,422 transactions within the time period described.<sup>5</sup> Divided by the respective total number of housing units (see Table 1), this indicates that only around 5% of the housing units had a transaction in the specified one-year period, suggesting that the transaction data are indeed very sparse. Figure 2 shows the total monthly transaction volume across all ten estates over time. Note that the monthly transaction volume shows significant month-to-month variations. In our dataset, the transaction volume is highest in February 2008, and has decreased to a lower level since then. Meanwhile, transaction volume is the lowest in August 2008 with only 274 transactions. The largest month-to-month change in transaction volume occurs between February 2008 and March 2008, with a decrease of 34.9%.

---

<sup>5</sup> Note that our sample size is somewhat smaller than that of comparable studies (e.g., Leung et al. 2002a, 2006, 2007). Furthermore, the ratio of total transactions to the total stock is below average in our data collection period. Further research may incorporate a larger sample size and a longer time horizon. We thank an anonymous reviewer for pointing out this issue.

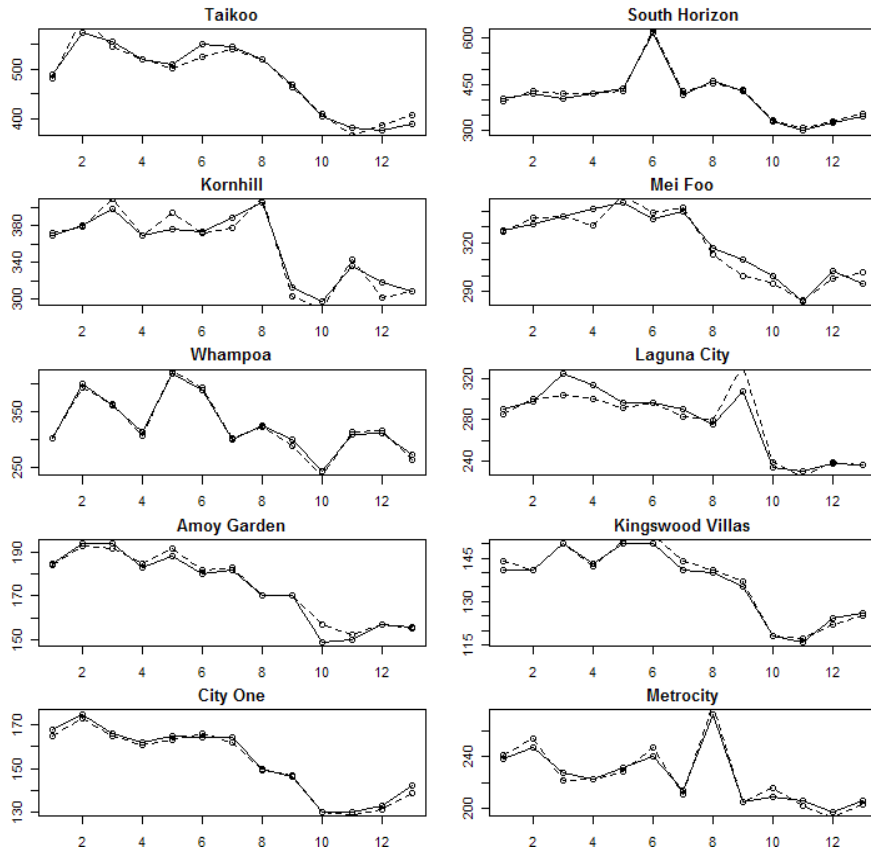
**Figure 2** Total Monthly Transaction Volume Across the Ten Estate Complexes in Our Dataset



Within each estate, transaction prices also exhibit significant variations over time, as shown in Figures 3 and 4, which display the median transaction prices of each of the ten estates over the one-year period, and the median per-square-foot transaction prices for each estate in each month.

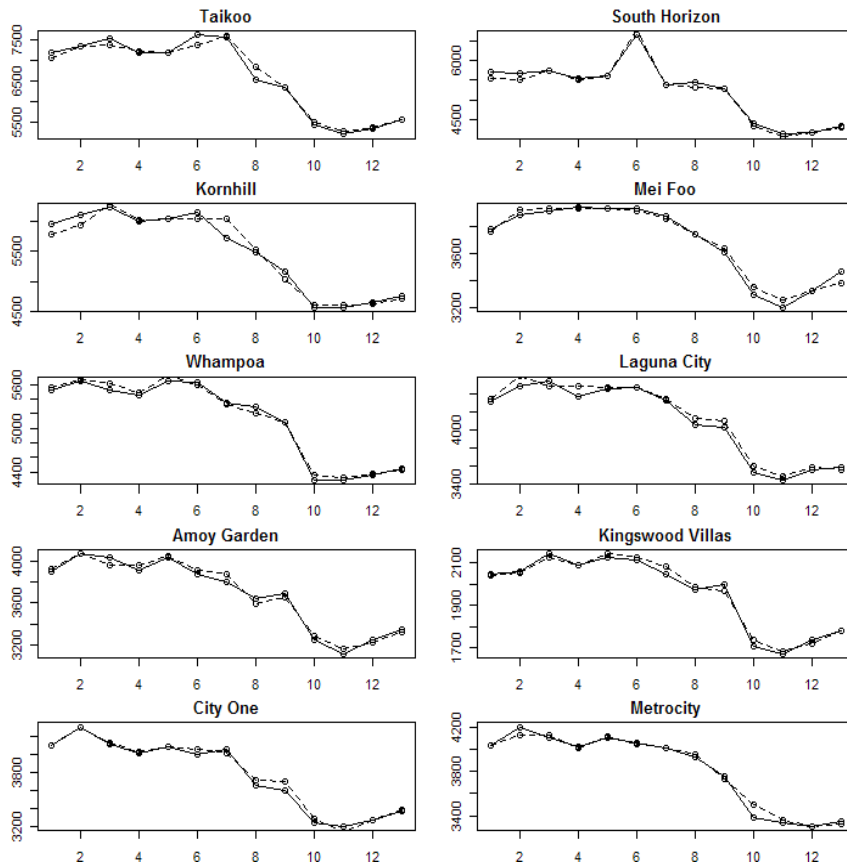
Since the median transaction price does not distinguish between housing units of different sizes and quality, the median price time series are very noisy and do not exhibit any clear pattern. The median per-square-foot price, on the other hand, shows a sharp decrease of around 12.7% (averaged across estates) from October 2008 to November 2008, presumably due to uncertainty and pessimism about the economy during the height of the global financial crisis. As evident in Figures 3 and 4, the median per-square-foot price varies from \$5,125 (in Hong Kong Dollars) to \$7,618 in the most expensive estate (Taikoo), and varies from \$1,673 to \$2,143 in the least expensive estate complex (Kingswood Villas).<sup>6</sup> After applying our model to the actual data, these figures will be used to validate our model to see if it preserves the observed patterns in transaction prices, both in absolute and in per-square-foot terms.

<sup>6</sup> Unless otherwise noted, all dollar signs in this paper refer to Hong Kong dollars, which is pegged to the U.S. dollar under a government guarantee (\$1 US = \$7.8 HK).

**Figure 3** Median Transaction Price for Each Estate in Each Month**Notes:****Solid Line:** The Observed Median Price.**Broken Line:** The Model-Fitted Median Price**Y-Axis:** Measured In \$10k (Hong Kong Dollar).**5. Empirical Results**

In this section, we calibrate our model to the transaction data described in Section 4. Using MCMC sampling, we obtain 500 samples from the posterior distribution of our model parameters. We discard the first 250 as “burn-in” samples (Gelman et al. 2003), and retain the remaining 250. In Section 5.1, we validate the fit of our model to in-sample data. In Section 5.2, we study and interpret the model parameters. In Section 5.3, we compare the out-of-sample valuation performance of our model against two benchmark methods.

**Figure 4 Median Per-Square-Foot Price (In Hong Kong Dollars) for Each Estate in Each Month**



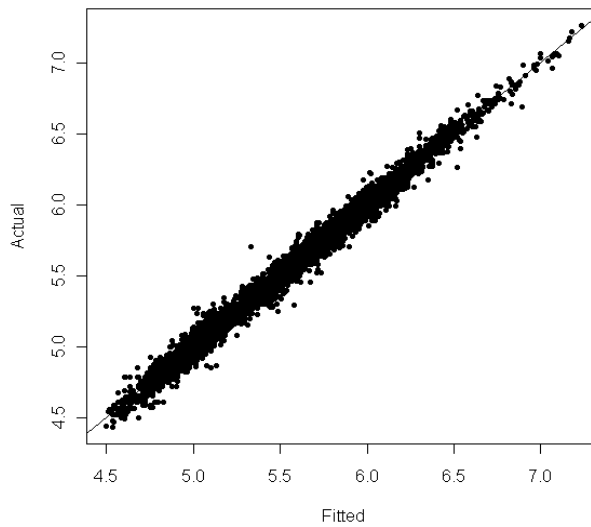
**Notes:**

**Solid Line:** The observed median per-square-foot price

**Broken Line:** The model-fitted median per-square-foot price

**5.1 Model Validation**

We validate the fit of our model on the summary statistics discussed in Section 4. Figure 5 plots the observed (log-) transaction prices against the model-fitted valuations for each of the estates in our dataset. As we can see, all the points on the scatterplot in Figure 5 lie close to the 45-degree line, showing that our model adequately describes the transaction data. Across all estates, the mean square error on log-scale is 0.0028; the mean absolute deviation is 0.040, again indicating an excellent in-sample fit.

**Figure 5** Actual versus Model-fitted (log-) Transaction Values

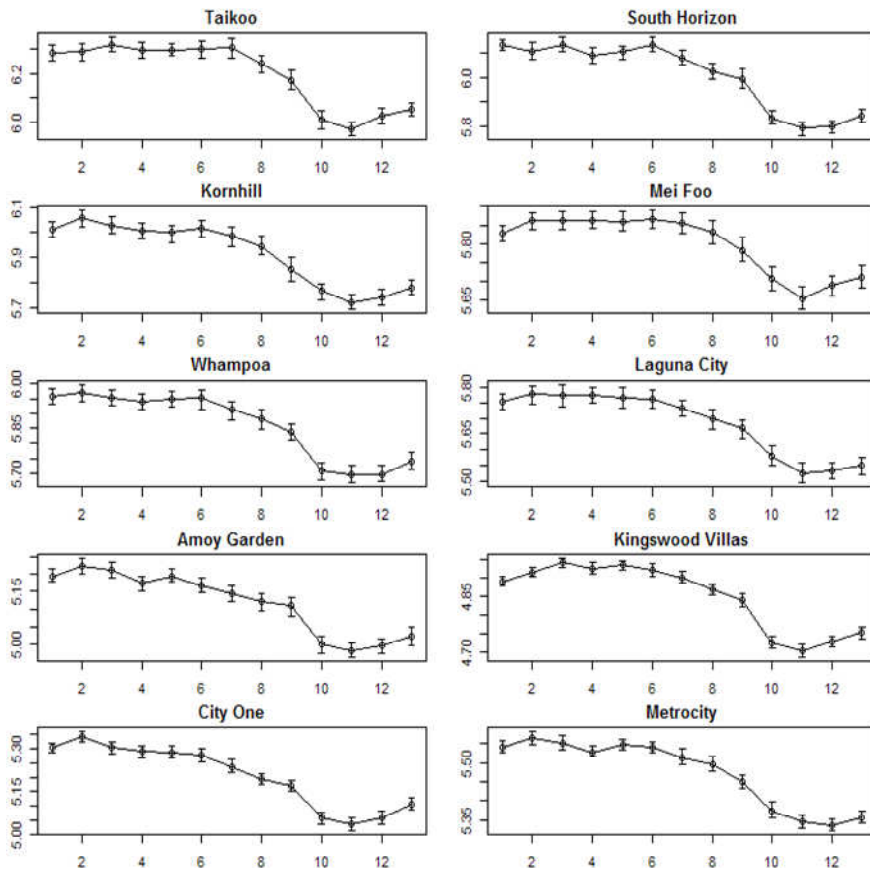
**Note:**  
**Solid Line:** 45-degree line

Figures 3 and 4 study the fit of our model on individual estates in terms of median transaction price and median per-square-foot price respectively. The solid lines in Figures 3 and 4 are the observed values, and the broken lines are the model-fitted values. As we can see in both figures, despite being parsimonious, our model is able to capture the idiosyncratic patterns in the price trends for each estate.

## 5.2 Parameter Interpretation

We now look at the parameter estimates from each model component in detail. We begin with the first model component, the “time effect” captured by  $\alpha_{it}$ , which is shown in Figure 6, along with the corresponding 95% posterior intervals (shown as vertical bars). As we discussed earlier,  $\alpha_{it}$  represents the general price levels of housing units within estate  $i$  at month  $t$ , and can be interpreted as a price index (Messe and Wallace 1997). By studying how  $\alpha_{it}$  changes over time, we can study how the general price levels of an estate are changing due to macroeconomic conditions. Since our data collection period (from February 2008 to February 2009) coincides with the onset and the height of the subprime financial crisis, the evolution of  $\alpha_{it}$  illustrates the effect of the financial crisis on the Hong Kong housing market.

**Figure 6** Posterior Mean of  $\alpha_{it}$  Plotted against Time (in Months) for Each Estate



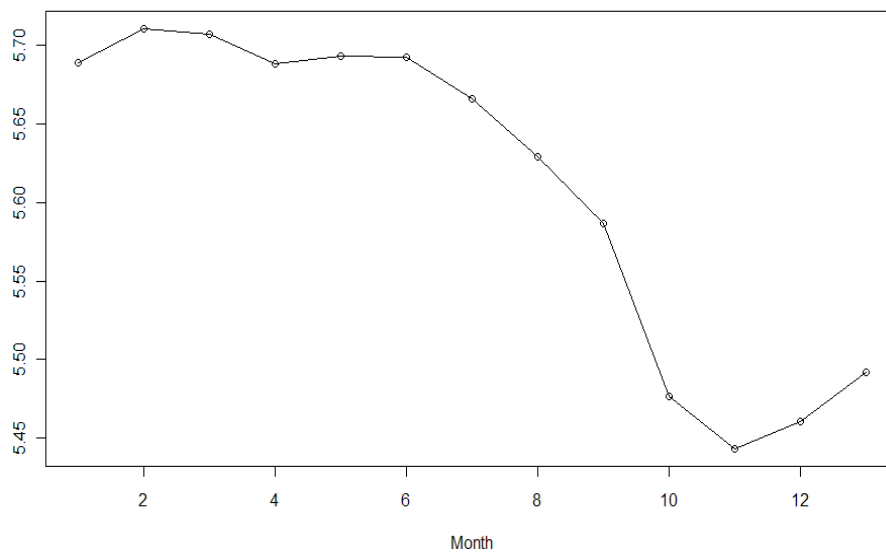
**Note:** The vertical bars denote the 95% posterior interval of  $\alpha_{it}$ .

From Figure 6, we see that the price levels ( $\alpha_{it}$ ) of all ten major estates are decreasing over time and show similar temporal patterns. In a comparison of the value of  $\alpha_{i(1)}$  and  $\alpha_{i(13)}$ , we find that the year-over-year changes in property value for the ten estates range from -0.115 (Mei Foo) to -0.287 (South Horizon). We also compute the average value of  $\alpha_{it}$  across all ten

estates (i.e.,  $\bar{\alpha}_t = \frac{1}{10} \sum_{i=1}^{10} \alpha_{it}$ ), and plot  $\bar{\alpha}_t$  over time in Figure 7. The year-

over-year change in the general price level ( $\bar{\alpha}_t$ ) across the ten estates is  $-0.198$ , which corresponds to a percentage decrease of  $1 - \exp(-0.198) = 18\%$  in the absolute price. In addition, we find that the drop in property price levels is most rapid from October 2008 to November 2008, which mirrors the pattern in the median price time series discussed in Section 4. Over that period, the general level of housing prices across the ten estates, drop by around 10.4%. However, by comparing this with the decrease of 12.7% (see Section 4) in the median price-per-square-foot level, we find that due to selection bias, the change in raw median price-per-square-foot overestimates the drop in general price levels. Thus, our model provides an index that automatically controls and corrects for any potential selection bias, allowing us to obtain a more accurate assessment of price changes.

**Figure 7** Plot of  $\bar{\alpha}_t$  (The General Price Pattern across Estates) Over Time



Next we turn our attention to the second and third model components, the block-unit effect  $\phi_{ij}$  and the floor effect  $\psi_{ijk}$ . For each block-unit in estate  $i$ , we have its size (in square feet) as the covariate ( $\bar{x}_{ij}$ ). Thus,  $\beta_i$  measures the sensitivity of property value to the size of the housing unit. For the floor effect,  $\gamma_i$  measures the floor premium (in log- scale) for estate  $i$ . Increasing the floor level by one increases the (log-) value by  $\gamma_i$  for a housing unit in estate  $i$ . The



posterior means and 95% posterior interval of  $\beta_i$  and  $\gamma_i$  for each estate are listed in Table 2.

**Table 2** Posterior Means and 95% Posterior Interval for  $\beta_i$  and  $\gamma_i$  Parameters for Each Estate

Estate name	Posterior mean for $\beta_i$ ( $\times 10^{-3}$ )	95% Posterior interval for $\beta_i$ ( $\times 10^{-3}$ )	Posterior mean for $\gamma_i$ ( $\times 10^{-3}$ )	95% Posterior interval for $\gamma_i$ ( $\times 10^{-3}$ )
<b>Taikoo</b>	1.75	(1.69, 1.82)	5.46	(4.44, 6.47)
<b>South Horizon</b>	1.99	(1.90, 2.08)	3.73	(3.01, 4.44)
<b>Kornhill</b>	1.76	(1.66, 1.85)	6.48	(5.32, 7.74)
<b>Mei Foo</b>	1.17	(1.13, 1.20)	5.66	(4.13, 7.20)
<b>Whampoa</b>	1.76	(1.70, 1.82)	8.16	(6.79, 9.82)
<b>Laguna City</b>	1.57	(1.48, 1.65)	5.87	(4.91, 6.90)
<b>Amoy Garden</b>	2.09	(1.81, 2.40)	3.90	(3.39, 4.39)
<b>Kingswood Villas</b>	1.44	(1.39, 1.50)	4.10	(3.77, 4.46)
<b>City One</b>	2.01	(1.92, 2.08)	4.17	(3.61, 4.78)
<b>Metrocity</b>	1.45	(1.41, 1.49)	2.66	(2.28, 3.04)

The results in Table 2 are generally in line with the previous literature, providing more face validity for the modeling approach. The posterior means of  $\beta_i$  and  $\gamma_i$  are all positive, with 95% posterior intervals that do not include zero. This is consistent with Choy et al. (2007) that a residential property of a larger size and a higher floor level commands a higher value ( $\beta_i > 0$  and  $\gamma_i > 0$ , respectively). The average value of  $\beta_i$  across the ten estates is 0.0017, indicating that increasing the size of a property by one square feet increases the (log-) property value by 0.0017 (0.17% in absolute scale) on average. The average value of  $\gamma_i$  across the ten estates is 0.0050, showing that increasing the floor level by one increases the (log-) property value by 0.0050 (0.5% in absolute scale) on average. As we can see in Table 2, there is a fair amount of variation in floor premium across the ten estates, ranging from 0.00266 (Metrocity) to 0.00816 (Whampoa). This variation can be explained by their difference in locations. For instance, estates built near the ocean might have a higher floor premium due to the better views on the top floors. In Section 6, we will discuss how our model for floor premium (Equation [4]) can be further generalized to a more flexible specification in future research.

### 5.3 Holdout Valuation

In order to study the ability of our model to provide accurate valuations, we performed two holdout assessments and compared the performance of our model to two competing benchmark methods commonly used by consumers in Hong Kong for property valuation. These two methods are the average price-per-sqft valuation, where consumers find out the price-per-sqft (psf) for the transactions in the same estate in the same month, compute the average psf, and multiply the average psf to the size of the unit of interest to infer value of the property, and expert assessment, where consumers visit the website of a major bank to obtain valuations that are made based on judgments of the real estate experts from the bank.

We performed two types of holdout assessment: (i) random, and (ii) chronological. For the random holdout assessment, we randomly selected 50 transactions from each estate as a holdout sample, which resulted in a holdout sample of a total of 500 transactions. Our model is then estimated based on the remaining transaction, and used to predict the valuation of the holdout units. We then compared the performance of our model against the valuation obtained by the average psf valuation method. The results are shown in Table 3, which demonstrates that our model consistently outperforms the average psf method in terms of holdout valuation. The mean absolute percentage error (MAPE) of our model is 6.00%, which represents a relative improvement of 25.1% over the MAPE of the average psf method (8.01%). The median absolute percentage error (MedAPE) of our model is 4.95%, which is a relative improvement of 25.8% when compared to the MedAPE of the average psf method (6.67%).

Note that because our holdout sample is randomly selected, our model carries the “advantage of hindsight”. That is, when making valuations for a transaction that occurred at time  $t_1$ , the model takes into account not only transactions that happened before  $t_1$ , but also transactions that occur after  $t_1$ . Thus, the “random” holdout assessment approach may be different from the actual situation where one has to make a valuation at time  $t$  with knowledge of transactions only up to time  $t$ .

To address this issue, we performed a “chronological” holdout assessment, where instead of selecting transactions randomly from the transaction dataset, we used the ten most recent transactions from each of the estates as a holdout sample, and calibrated our model with the remaining data. Thus, in this assessment, our model makes valuation-based only transactions that occur before the focal transaction as is the case in reality. We compared our model’s performance with: (i) the average psf valuation approach, and (ii) the valuation based on a bank expert assessment. The bank’s valuation was obtained through an online interface of a major bank in Hong Kong.

**Table 3 Holdout Valuation Results of Our Model, Compared To Average Per-Square-Foot Price Based Valuation**

Estate name	Model-based valuation		Average price-per-square-foot valuation	
	Mean Abs. % Error (MAPE)	Median Abs. % Error (MedAPE)	Mean Abs. % Error (MAPE)	Median Abs. % Error (MedAPE)
Tai Koo	7.83	6.68	11.01	9.43
South Horizon	5.50	4.31	9.73	8.01
Kornhill	5.98	5.98	7.47	6.32
Mei Foo	7.58	6.23	8.17	6.20
Whampoa	5.72	4.41	8.27	8.57
Laguna City	5.33	4.45	7.17	5.76
Amoy Garden	4.77	3.41	6.03	5.33
Kingswood Villa	5.96	5.14	7.35	6.09
City One	7.29	5.26	9.55	6.38
Metrocity	4.02	3.61	5.35	4.61
Overall	6.00	4.95	8.01	6.67

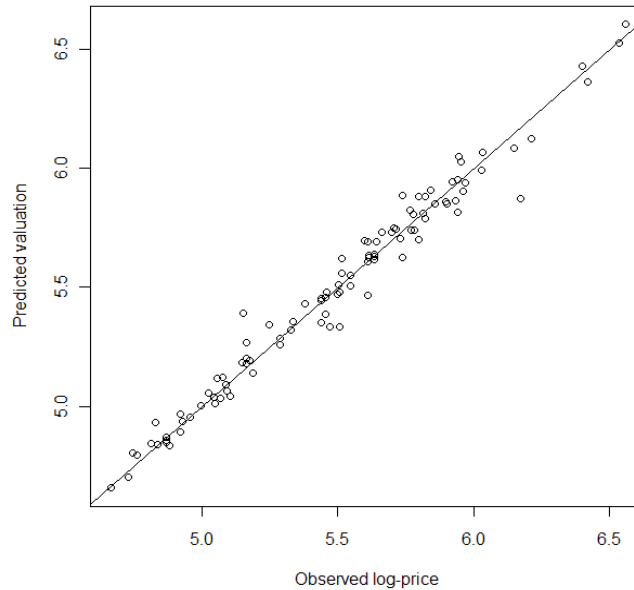
**Note:** The holdout sample is comprised of 500 transactions randomly drawn from our transaction dataset (50 transactions in each estate).

**Table 4 Holdout Valuation Results (With the Holdout Sample Defined As the Ten Most Recent Transactions in Each Estate), Compared To Two Competing Benchmark Methods Based on (I) Average Per-Square-Foot Price and (ii) Expert Assessment by a Major Bank in Hong Kong**

Estate name	Model-based valuation		Average per-square-foot valuation		Bank expert assessment	
	Mean Abs % Error	Median Abs % Error	Mean Abs % Error	Median Abs % Error	Mean Abs % Error	Median Abs % Error
Tai Koo	6.24	5.60	16.81	15.62	4.98	5.25
South Horizon	3.03	2.96	12.16	8.80	4.71	3.45
Kornhill	6.17	4.63	6.08	3.47	6.02	5.42
Mei Foo	6.57	5.14	9.64	7.76	12.07	10.21
Whampoa	4.42	2.82	5.47	5.61	3.67	3.24
Laguna City	5.88	5.05	5.13	6.21	7.51	7.11
Amoy Garden	3.55	3.25	3.94	3.47	4.98	4.33
Kingswood Villa	2.61	2.76	4.78	3.60	3.98	2.48
City One	7.59	3.30	10.30	10.49	8.58	6.78
Metrocity	2.36	2.07	3.04	2.87	2.00	1.48
Overall	4.84	3.41	7.74	5.43	5.85	4.62

The summary results of the chronological holdout assessment are shown in Table 4. Figure 8 plots the valuation of our model versus the actual (log-) transaction prices. The figure shows that our model is able to assess the value of each housing unit and hence, predict the transaction price with small errors. Table 4 shows that our model consistently outperforms both benchmark methods. In particular, the MAPE of our model is 4.84%, which compares favorably with that of average psf valuation (7.74%) and expert assessment (5.85%), across the ten estates. We conclude that both studies suggest that our model provides superior valuation performance in comparison to the benchmark method widely used by consumers in Hong Kong.

**Figure 8** Plot of (Log-) Observed versus Predicted Transaction Price in Out-Of-Sample Assessment



## 6. Conclusion and Future Research

In this paper, we have developed a hierarchical Bayesian approach to model property value in Hong Kong, taking into account the unique structure of the Hong Kong residential real estate market. Our model consists of three components: (i) time effect, (ii) block-unit effect, and (iii) floor effect, which are combined with a stochastic error term. We have calibrated our model on data that cover all transactions in ten major residential estates in Hong Kong from February 2008 to February 2009. Based on estimates from our model,

we find that the year-to-year change in overall property price levels during the data collection period is about -18%. The largest month-to-month change occur from October 2008 to November 2008, with a drop of over 10% across all property values, which reflects the impact of the global financial crisis on the Hong Kong housing market over that time period.

Furthermore, we have explored the performance of our model in valuating properties based on past transactions through two sets of holdout assessments. We find that our model outperforms: (i) the average psf valuation, and (ii) the expert assessment from a major bank in Hong Kong, which are two of the most common sources of property evaluation data for Hong Kong consumers. Given that our model provides significant value for consumers, we plan to publish our model-based valuations online at no cost, with the goal of improving transparency in the Hong Kong housing market and allowing consumers to make more informed investment decisions. Our model may also be useful for banks that are interested in obtaining an accurate risk assessment of their mortgage loan portfolios.

Given that our proposed model is a Bayesian extension to hedonic pricing models, it suffers from the same kinds of limitations (although to a smaller extent), as discussed in Lentz and Wang (1998). First, if there are too few transactions in an estate/block unit, the resulting standard errors and uncertainty around valuations may be too large to be acceptable. In such cases, one may need to incorporate more shrinkage, using a tighter hyper-prior distribution (Gelman et al. 2003) across estates/block units in order to reduce posterior uncertainty. Second, it is unclear what types of unit features should be included in our valuation model; statistically, this is a variable-selection problem that can be addressed using the Gibbs sampling approach proposed by George and McCulloch (1993). Finally, in this article we do not consider transactional variables (“conditional of sale” variable) that may also affect valuations (Lentz and Wang 1998); e.g., units that are sold under distressed or adverse financing conditions may command a lower price. The ways that transactional variables and other “outlier” observations should be identified and adjusted within our model framework may be addressed in future research.

Taking this research one step further, future studies may consider extending our model framework both in terms of other housing unit covariates and a more flexible model specification. We briefly discuss these topics below:

(i) Incorporate additional unit characteristics: Due to the limited information in our dataset, in this work, we only included the size (in sqft) of housing units as covariates. Future research may consider including more covariates, some of which have been identified in previous research that affect property value, into Equation [3] of our model. These might include quality of the view (mountain, park, cemetery, ocean view) of the unit, the directionality, amenities, and closeness to public transport (e.g., Bao and Wan 2004, Choy et

al. 2007). Incorporating a richer set of unit characteristics into the model may further improve valuation performance.

(ii) Generalize model specifications: As mentioned in Section 3, the linear specification of the floor effect can be relaxed to allow a non-linear floor effect, where some evidence of such is provided by Choy et al. (2007). In particular, instead of specifying that the floor effect is a linear function of floor level as stated in Equation (4), we can model the floor premium of each floor separately, and link them together through a hyperdistribution. Formally, we can generalize Equation [4] as follows:

$$\psi_{ijk} = \psi_{ij(k-1)} + \kappa_{ijk} \quad \kappa_{ijk} \sim N(\mu_{\kappa}, \sigma_{\kappa}^2) \quad (4^*)$$

where  $\kappa_{ijk}$  denotes the floor premium of the  $k$ -th floor in estate  $i$ , block-unit  $j$ . This allow for a non-parametric way to the model floor effect.

(iii) Time-varying parameter(s): Similarly, we can relax our model assumption on  $\beta_i$  to allow for the possibility that it can be time-varying. For instance, Leung et al. (2002a, b, 2006, 2007) show that the pricing of different housing attributes can change over time; as a result, biased estimation may result if a researcher pools all the data into one regression and only uses a time-dummy to control for time variation.<sup>7</sup> To incorporate the time-varying parameters within our model framework, we can generalize Equation (3) to (3\*) as follows:

$$\phi_{ijt} = \bar{x}_{ij}' \beta_{it} + \delta_{ijt}, \quad (3^*)$$

(iv) Comparison to other benchmarks: In this article, we have compared the performance of our model to two benchmarks that are commonly used in market practice. In future research, one may also further assess the performance of our model framework by comparing it against other models in the economics literature, such as GARCH-type and ARMA-type models (e.g., Brockwell and Davis 2003), or other methods that are recently proposed in the real estate literatures (e.g., the replication method proposed by Lai et al. 2008; and the “adjustment grid” method used in Vandell 1991, and Lai and Wang 1996).

To conclude, we believe that this paper provides the first step in applying hierarchical Bayesian models for objective valuations in the Hong Kong property market. Future work can build upon and further extend our model to improve property valuations, which ultimately benefits both consumers and the bank industry by increasing market transparency.

---

<sup>7</sup> We thank an anonymous reviewer for pointing out this issue.

## Reference

- Bailey, Martin J., Richard F. Muth, and Hugh O. Nourse. (1963). A Regression Model for Real Estate Price Index Construction, *Journal of the American Statistical Association*, **58**, Dec, 933-942.
- Ball, M. (1973). Recent Empirical Work of the Determinants of Relative House Prices, *Urban Studies*, **10**, 213-233.
- Bao, Helen X.H., and Alan T.K. Wan. (2004). On the Use of Spline Smoothing in Estimating Hedonic Housing Price Models: Empirical Evidence Using Hong Kong Data, *Real Estate Economics*, **32**, 3, 487-507.
- Brockwell, Peter J., and Richard A. Davis. (2003). *Introduction to Time Series and Forecasting*, Springer, New York.
- Case, Karl E., and Robert J. Shiller. (1989). The Efficiency of the Market for Single-Family Homes, *American Economic Review*, **79**, 1, 125-137.
- Casella, George, and Edward George. (1992). Explaining the Gibbs Sampler, *The American Statistician*, **46**, 3, 167-174.
- Chan, L.K., Y.C. Chan, Y.V. Hui, H.P. Lo, S.K. Tse, A. Wan, and K. Yau. (2009). Centa-City Index: Hong Kong's Definitive Property Price Indices, *Technical Note*, available at: <http://fbweb.cityu.edu.hk/ms/cci/content.htm>.
- Chau, K.W. (2006). The University of Hong Kong Real Estate Index Series (HKU-REIS): Index Construction Method for The University of Hong Kong All Residential Price Index (HKU-ARPI) and the following Sub-Regional Residential Price Indices, *Technical Note*, available at [http://hkureis.versitech.hku.hk/HKU\\_Index-description\\_Ver1\\_28\\_.pdf](http://hkureis.versitech.hku.hk/HKU_Index-description_Ver1_28_.pdf).
- Chau, K.W., V. S. M. Ma, and D. C. W. Ho. (2001). The Pricing of 'luckiness' in the Apartment Marketing, *Journal of Real Estate Literature*, **9**, 1, 31-40.
- Chau, K.W., K.S.K. Wong, and E.C.Y. Yiu. (2004). The Value of the Provision of a Balcony in Apartment in Hong Kong, *Property Management*, **22**, 3, 250-264.
- Choi, Jeonghye, Sam K. Hui, and David Bell. (2009). Bayesian Spatio-Temporal Analysis of Imitation Behavior in Adoption of an Online Grocery Retailer, *Journal of Marketing Research*, forthcoming.

Choy, Lennon H.T., Stephen W. K. Mak, and Winky K. O. Ho. (2007). Modeling Hong Kong Real Estate Prices, *Journal of Housing and Built Environment*, **22**, 359-368.

Clapham, Eric, Peter Englund, and John M. Quigley. (2006). Revisiting the Past and Settling the Score: Index Revision for House Price Derivatives, *Real Estate Economics*, **34**, 2, 275-302.

Cochrane, John H. (1988). How Big is the Random Walk in GNP? *The Journal of Political Economy*, **96**, 5, 893-920.

Cropper, Maureen L., Leland B. Deck, and Kenneth E. McConnell. (1988). On the Choice of Functional Form for Hedonic Price Functions, *The Review of Economics and Statistics*, **70**, 4, 668-675.

Engel, C.M., and J.D. Hamilton. (1990). Long Swings in the Dollar: Are They in the Data and Do Markets Know It?, *American Economic Review*, **80**, 689-713.

Freeman, A. M. (1979). Hedonic Prices, Property Values, and Measuring Environmental Benefits: A Survey of the Issues, *Scandinavian Journal of Economics*, **81**, 154-171.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. (2003). *Bayesian Data Analysis, 2<sup>nd</sup> Edition*, Chapman & Hall.

George, Edward I., and Robert E. McCulloch. (1993). Variable Selection Via Gibbs Sampling, *Journal of the American Statistical Association*, **88**, 423, 881-889.

Gillard, Q. (1981). The Effect of Environment Amenities on House Values: The Example of a View Lot, *Professional Geographer*, **33**, 216-220.

Greene, William H. (2007). *Econometric Analysis, 7<sup>th</sup> Edition*, Prentice Hall.

Hui, Eddie, and Shen Yue. (2006). Housing Price Bubbles in Hong Kong, Beijing, and Shanghai: A Comparative Study, *Journal of Real Estate Finance and Economics*, **33**, 4, 299-327.

Lai, Tsong-Yue, Kerry Vandell, Ko Wang, and Gerd Welke. (2008). Estimating Property Values by Replication: An Alternative to the Traditional Grid and Regression Methods, *Journal of Real Estate Research*, **30**, 4, 441-460.



- Lai, Tseong-Yue, and Ko Wang. (1996). Comparing the Accuracy of the Minimum-Variance Grid Method to Multiple Regression in Appraised Value Estimates, *Real Estate Economics*, **24**, 531-549.
- Lancaster, K. J. A. (1966). A New Approach to Consumer Theory, *Journal of Political Economy*, **74**, 132-157.
- Leggett, C.G., and N.E. Bockstael. (2000). Evidence of the Effects of Water Quality on Residential Land Prices, *Journal of Economics and Management*, **39**, 121-144.
- Lentz, George H., and Ko Wang. (1998). Residential Appraisal and the Lending Process: A Survey of Issues, *Journal of Real Estate Research*, **15**, 1/2, 11-39.
- Leung, C. K. Y., G. C. K. Lau, and Y. C. F. Leong. (2002a). Testing Alternative Theories of the Property Price-Trading Volume Correlation, *Journal of Real Estate Research*, **23**, 3, 253-263.
- Leung, C. K. Y., Y. C. F. Leong, and I. Y. S. Chan. (2002b). TOM: Why Isn't Price Enough? *International Real Estate Review*, **5**, 1, 91-115.
- Leung, C. K. Y., Y. C. F. Leong, and S. K. Wong. (2006). Housing Price Dispersion: An Empirical Investigation, *Journal of Real Estate Finance and Economics*, **32**, 3, 357-385.
- Leung, C. K. Y., S. K. Wong, and P. W. Y. Cheung. (2007). On the Stability of the Implicit Prices of Housing Attributes: A Dynamic Theory and Some Evidence, *International Real Estate Review*, **10**, 2, 65-91.
- Li, M.W. and H.J. Brown. (1980). Micro-neighbourhood Externalities and Hedonic Housing Prices, *Land Economics*, **56**, 2, 125-141.
- Mankiw, N. Gregory, and Matthew D. Shapiro. (1985). Trends, Random Walks, and Tests of the Permanent Income Hypothesis, *Journal of Monetary Economics*, **16**, 165-174.
- Messe, Richard A., and Nancy E. Wallace. (1997), The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic-Regression, and Hybrid Approaches, *Journal of Real Estate Finance and Economics*, **14**, 51-73.
- Mok, H.M.K., P.P.K. Chang, and Y.S. Cho. (1995). A Hedonic Price Model for Private Properties in Hong Kong, *Journal of Real Estate Finance and Economics*, **10**, 1, 37-48.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *Journal of Political Economy*, **82**, 1, 35-55.

Sun, Hua, Yong Yu, and Shi-Ming Yu. (2005). A Spatio-Temporal Autoregressive Model for Multi-Unit Residential Market Analysis, *Journal of Real Estate Finance and Economics*, **31**, 2, 155-187.

Tiebout, C. M. (1956). A Pure Theory of Local Expenditure, *Journal of Political Economy*, **65**, 416-424.

Vandell, K. D. (1991). Optimal Comparable Selection and Weighting in Real Property Valuation, *Journal of the American Real Estate and Urban Economics Association*, **19**, 213-39.

Zhou, Sherry Z., and Helen X. Bao. (2007). Modelling Price Dynamics in the Hong Kong Property Market, *Working paper*, available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1012320](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1012320).

## Appendix I: MCMC Sampling Procedure

As we discussed in Section 3.2, all our model parameters are given conjugate, weakly informative prior distributions (Equations (5) – (9)). With these conjugate priors, the full conditional distribution for all parameters is of standard form. The Gibbs sampler (Casella and George 1992) is used to sample from them. In the discussion below, we outline the full conditional distribution for all model parameters. The following three results (see Choi et al. 2009, Gelman et al. 2003) are used:

(i) First, consider a vector of  $y$  of  $n$  i.i.d. observations from a normal distribution with mean  $\mu$  and known variance  $\sigma^2$ . Given a conjugate prior on  $\mu$ , in the form  $\mu \sim N(\mu_0, \sigma_0^2)$ , the full conditional distribution of  $\mu$  is:

$$\mu | y \sim N\left(\frac{(\sigma_0^2)^{-1} \mu_0 + (\sigma^2/n)^{-1} \bar{y}}{(\sigma_0^2)^{-1} + (\sigma^2/n)^{-1}}, \frac{1}{(\sigma_0^2)^{-1} + (\sigma^2/n)^{-1}}\right) \quad [\text{A-1}]$$

(ii) Second, consider a linear model  $y | \vec{\beta}, \sigma^2, \bar{x} \sim N(\bar{x}' \vec{\beta}, \sigma^2)$  with known variance  $\sigma^2$ . Given the conjugate prior on  $\vec{\beta}$  (in the form  $\vec{\beta} \sim N(\beta_0, \sigma_\beta^2 I)$ ), the full conditional distribution is:

$$\vec{\beta} | y \sim N\left(\left((\sigma^2)^{-1} X'X + (\sigma_\beta^2)^{-1} I\right)^{-1} \left((\sigma^2)^{-1} X'y + (\sigma_\beta^2)^{-1} \beta_0\right), \left((\sigma^2)^{-1} X'X + (\sigma_\beta^2)^{-1} I\right)^{-1}\right) \quad [\text{A-2}]$$

(iii) Third, consider a vector  $y$  of  $n$  i.i.d. observation from a normal distribution with known mean  $\mu$  and unknown variance  $\sigma^2$ . Given a conjugate prior on  $\sigma^2$  (in the form  $\sigma^2 \sim \text{Inv} - \chi^2(v_0, s_0^2)$ ), the full conditional distribution is:

$$\sigma^2 | y \sim \text{Inv} - \chi^2\left(v_0 + n, \frac{v_0 s_0^2 + n s^2}{v_0 + n}\right),$$

$$\text{where } s^2 = \frac{\sum (y_i - \mu_i)^2}{n} \quad [\text{A-3}]$$

We now outline the ways that we sample each individual model parameter.

### Time effect $\alpha_{it}$

Let  $\theta_{it} = y_{ijkt} - \phi_{ij} - \psi_{ijk}$ . Then  $\theta_{it} | \alpha_{it}, \sigma_y^2 \sim N(\alpha_{it}, \sigma_y^2)$ . Priors for  $\alpha_{it}$  are given below, depending on the time periods:

$$\begin{aligned}
\text{For } t = 1, \quad & \alpha_{it} \sim N\left(\frac{(\sigma_\alpha^2)^{-1}(\alpha_{i2} - \mu_i)}{(\mathbf{100}^2)^{-1} + (\sigma_\alpha^2)^{-1}}, \frac{1}{(\mathbf{100}^2)^{-1} + (\sigma_\alpha^2)^{-1}}\right) \\
\text{For } 1 < t < T, \quad & \alpha_{it} \sim N\left(\frac{\alpha_{i(t-1)} + \alpha_{i(t+1)}}{2}, \frac{\sigma_\alpha^2}{2}\right) \quad [\text{A-4}] \\
\text{For } t = T, \quad & \alpha_{it} \sim N(\alpha_{i(T-1)} + \mu, \sigma_\alpha^2)
\end{aligned}$$

With this prior distribution, we apply Equation [A-1] to sample  $\alpha_{it}$ .

#### Block-unit effects $\phi_{ij}$

Let  $\theta_{it} = y_{ijkt} - \alpha_{it} - \psi_{ijk}$ . Then,  $\theta_{it} | \phi_{ij} \sim N(\phi_{ij}, \sigma_y^2)$ . With the prior  $\phi_{ij} \sim N(x_{ij}' \beta_i, \sigma_\phi^2)$ , we apply Equation [A-1].

#### Floor effect $\phi_{ij}$

Let  $\theta_{it} = \frac{y_{ijkt} - \alpha_{it} - \phi_{ij}}{k}$ . Then  $\theta_{it} | \gamma_i \sim N(\gamma_i, \frac{\sigma_y^2}{k^2})$ . With the prior  $\gamma_i \sim N(0, \mathbf{100}^2)$ , we apply Equation [A-1].

#### Block-unit covariate effect $\beta_i$

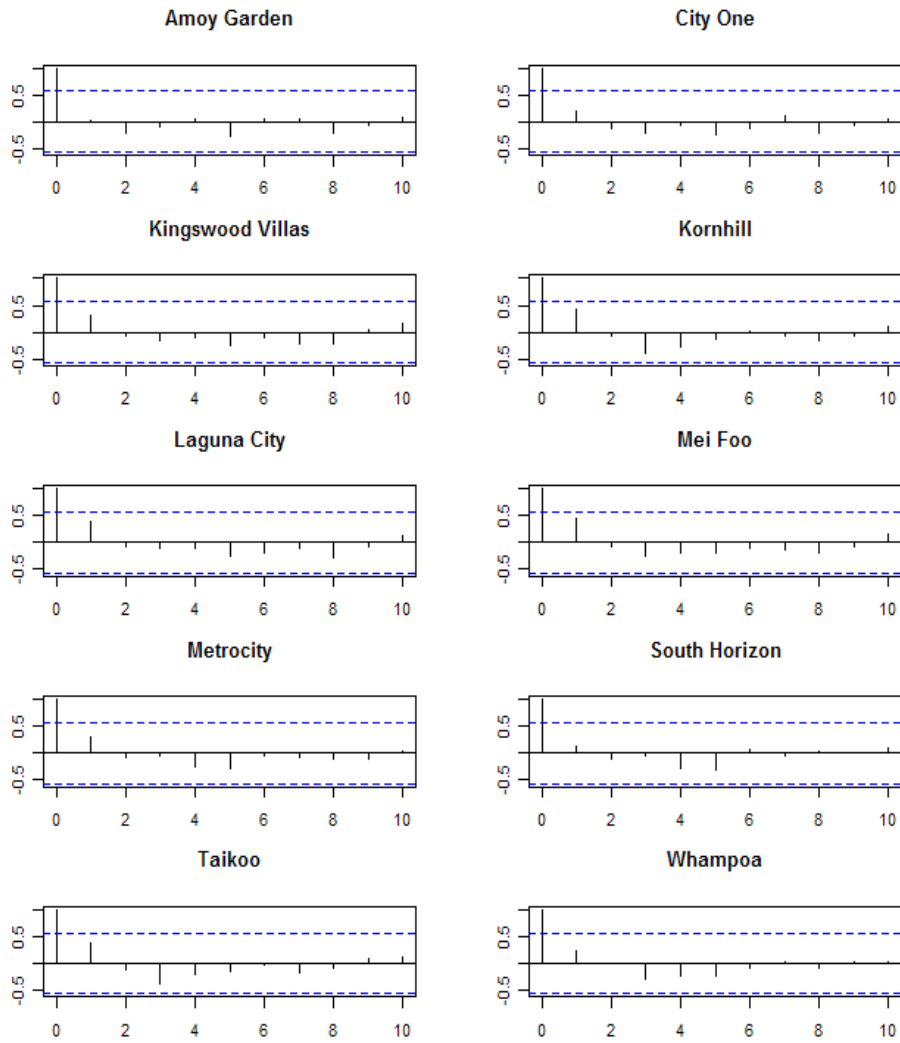
We have  $\phi_{ij} | \beta_i \sim N(x_{ij}' \beta_i, \sigma_\phi^2)$ . With the prior  $\beta_i \sim N(0, \mathbf{100}^2 I)$ , we apply Equation [A-2] to sample from the posterior distribution of  $\beta_i$ .

#### Variance parameters $\sigma_{(i)y}^2, \sigma_{(i)\phi}^2, \sigma_{(i)\alpha}^2$

The variance parameters are all sampled in the same manner using Equation [A-3]. For  $\sigma_{(i)y}^2$ , let  $\theta_{it} = y_{ijkt} - \alpha_{it} - \phi_{ij} - \psi_{ijk}$ , hence  $\theta_{it} \sim N(0, \sigma_{(i)y}^2)$ .

With conjugate prior on  $\sigma_{(i)y}^2$ , Equation [A-3] is used.

**Appendix II: Autocorrelation Plots of  $W_{it}$**



As evident from the above autocorrelation plots, all of the autocorrelations (of different lags) are below the broken line, which indicates that none of them are statistically significant. This provides some empirical support that  $Cov(w_{it}, w_{i,t-n}) = 0$ , one of the assumptions behind our random-walk-with-drift specification.