INTERNATIONAL REAL ESTATE REVIEW

A Localized Model for Residential Property Valuation: Nearest Neighbor with Attribute Differences

Simon K.C. Cheung
Department of Statistics, Room 119, Lady Shaw Building, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China. Phone: 852-27686811. Email: simoncheungkc@gmail.com

A special type of spatial dependence model has been developed for the valuation of residential property by using a Bayesian approach which takes into account the unique attributes of Hong Kong property. Inspired by an approach that bases transactions on nearest neighbor estimation, the current model has fine-tuned the spatial dependence method of deriving value from the transactions of the real estate closest in proximity by factoring in differences in structural and auxiliary attributes for more precise property valuation. The model is established on a dataset that covers all residential real estate transactions in a new town in Hong Kong during a given period of time. This model excels other valuation methods with its unique reference to attribute differences, which may impact the valuation results of properties in close vicinity. With this calibration method, we hope to improve the accuracy of real estate appraisals, thus enabling potential homeowners, lenders and investors to make better informed decisions.

Keywords

Prediction Accuracy, Nearest Neighbor Model, Attribute Differences, Hedonic Price Model, Hong Kong Housing, Forecastability

1. Introduction

Among the different kinds of assets, equity represents a significant portion of the investment portfolio of most households in Hong Kong. It is estimated to take up on average about 40% of the household net worth (Gelfand et al. 1998). An accurate valuation based on the performance of the real estate market and the unique attributes of a property is therefore important for many stakeholders, including property buyers, investors, and financial institutions. For buyers, an exact assessment of the potential purchase is crucial for making financial arrangements on the amount of down payment, size of the mortgage and repayment period. The valuation result may sometimes give the buyer and real estate agent some grounds to negotiate with the seller on a probable and realistic selling price. For financial institutions, a more refined valuation method facilitates better risk calculation and well-informed lending decisions on loan applications. Banks use the appraisal results to recommend suitable mortgage products that bring probable and profitable businesses to them (Isaac and O'Leary 2012).

This paper proposes a property valuation method that combines the merits of the nearest neighbor approach and the measurement of attribute differences between a property and its neighbors to arrive at a more precise estimation of house value. This special spatial dependence model that takes into account attributes at the both macro and micro levels originates from the following observations.

First, the structure of some of the Hong Kong private residential estates built in recent years differs largely from those built twenty to thirty years ago. In the past, most high-rise condominiums were built by designing units that share the same floor plan within the same building or a housing estate made up of multiple condominium buildings that replicate a common design (Hui and Wong 2009). In the last two decades, developers have been increasingly partnering up to bid for larger development sites to allow for the development of well-planned estates with upmarket and value-added auxiliary features, for example, clubhouses with playrooms, indoor gymnasiums and bowling alleys, heated pools, and barbecue sites (Lam 2008) in order to increase the sales value and the recurrent income from property management. To cater to property buyers who share the same social and recreational interests, but different purchasing power, developers diversify the structural design of units across each floor, within the same building and across the whole housing estate. As will be discussed in Section 2, these newly developed estates are characterized by a mix of units of different sizes, configurations and features. They may range from small-sized units (less than 500 square feet with 1-2 rooms and 1 bathroom), large-sized units (larger than 1,000 square feet with balcony and 2-3 bathrooms) to a small number of feature units (duplex, adjoining, with private rooftop, pool or garden). The presence of attribute differences, structural or auxiliary, is found to have varying impacts on property value, which may not be easily captured by the framework of the nearest neighbor approach. To offer a fine-tuned model to account for attributes at both the macro and micro levels, a valuation approach based on a nearest neighbor model with adjustment of attribute differences is thus essential.

Second, attribute differences are also a uniquely crucial factor for consideration in the valuation of some of the less recent Hong Kong private residential properties that comprise high-rise condominiums which share similar floor plans but differ in attributive variables e.g. directionality, quality of view (sea, mountain, garden, cemetery), and accessibility to public transport (Hui and Wong 2009). For a more precise valuation of real estate in a densely packed vertical city like Hong Kong, it is essential to consider a multitude of salient attributes, including structural attributes (e.g. floor area, ceiling height, number and size of rooms and balcony), locational attributes (e.g. floor level, facing view, proximity to city centre and public transportation), and auxiliary attributes (e.g. clubhouse, shopping mall and carpark). Chau (2005) found that the balcony has a positive effect on the value of a Hong Kong property, irrespective of the quality of the view. This finding is in line with the importance of good design and availability of a clubhouse in home buying decisions (Lam 2008). By applying a linear-in-means approach (Manski 1993) to provide an initial valuation, the current model further modifies the estimated value with attribute differences between a property and its neighbors to calculate precise value differences between units of a heterogeneous nature in a close neighborhood. Given the similarity between the Hong Kong real estate market and those of other Asian cities with comparable macro- and micro-level differences, the present model is likely to be applicable to other real estate markets.

Third, to test the performance of the statistical model in this study and compare the result with that of other valuation methods based on spatial dependence, a dataset on all the recorded residential transactions in a 3-month period of 2013 in a new town in Hong Kong (described in detail in Section 4) is chosen for two reasons. First, this new town has been developed as a satellite residential district since the mid-1980s. An array of residential properties with different features, both structural and auxiliary, are found and the transactions that occurred reflect the importance of attribute differences in property value. Second, the data collection period falls in the recovery period of the global subprime crisis in 2008 and before the overheating of the market from the second quarter of 2014. The buying market was rational and allows our model parameters to provide truthful insights on the effects of attribute differences on Hong Kong real estate prices in an effective market.

Throughout the literature, previous valuation models suffer from four limitations. These limitations are summarized by Caplin et al. (2008) and Hui and Wong (2009). First, prior approaches predict house prices by referring to the transactions of the same residential units in the recent past. Such prediction by lag of prices could be problematic since transactions of the same units tend to be “sparse” and may not form a reliable basis for forecasting changes in future housing prices (Caplin et al. 2008). Second, some previous pricing models require an exhaustive list of regressors for generating a meaningful function. However, the data on all relevant attributes of a residential unit and its neighborhood may either never be recorded or too expensive to obtain (Can 1992; Caplin et al. 2008). Hence, omitted features could be a fundamental problem in linear parametric pricing functions. Third, it is observed that some pricing models such as the spatial dependence model (Elhorst 2010, Caplin et al. 2008) have an inherent issue of information overlap when the transactions of the nearest neighbors of an apartment have already taken into account some unspecified features, which may impact the effectiveness of the feature list and alter the estimation results. Finally, the imposition of the restriction on the spatial parameter $\rho < 1$ to ensure the invertibility of the matrix $I - \rho W$, where W is a row-normalized matrix of the spatial arrangement of the spatial units in the sample and I is an identity matrix (Elhorst 2010), is actually less intuitive to real estate valuation than by a mere average of the transactions of its nearest neighbors. Therefore, in order to estimate the price difference of units in close proximity, a more appropriate pricing model should take into account attribute differences.

The paper contributes to the literature in the following ways. First, to the best of our knowledge, this is the first paper that uses the nearest neighbor approach with adjustment of attribute differences between residential units in enhancing the accuracy and efficiency of the valuation of real estate. This modified approach addresses two methodological challenges: (i) the nearest neighbor approach resolves the problem of infrequent data on the transaction of same units and accounts for the changes in macro-level variables such as locational, physical and economic variables, which strongly influence the market price as reflected in the recent transacted price (Can 1990); and (ii) the extension of the approach with adjustment of attribute differences offers a solution to the often imperfect and incomplete list of variables, which is a fundamental issue inherent in hedonic pricing and conventional spatial dependence models (Caplin et al. 2008). Our proposed model mitigates this inadequacy of hedonic and spatial dependence models and extends the nearest neighbor approach by acknowledging the importance of micro-level variable differences in the valuation of residential units in proximity of each other. This modified approach offers an easy and accurate two-step tool to estimate composite home value by means of spatial autocorrelation and comparison of attribute differences, which makes it outperform hedonic, conventional spatial dependence or expert assessment-based valuation methods.

This paper is organized into five parts. Section 2 presents information on the Hong Kong property market, and provides a brief review of previous models on real estate valuation. Section 3 is a description on the conventional spatial dependence model and our improved nearest neighbor model that takes into account attribute differences. Section 4 is a description of our dataset. Section 5 reports our empirical results from estimating a value prediction function based on a Bayesian approach. Finally, Section 6 concludes the study.

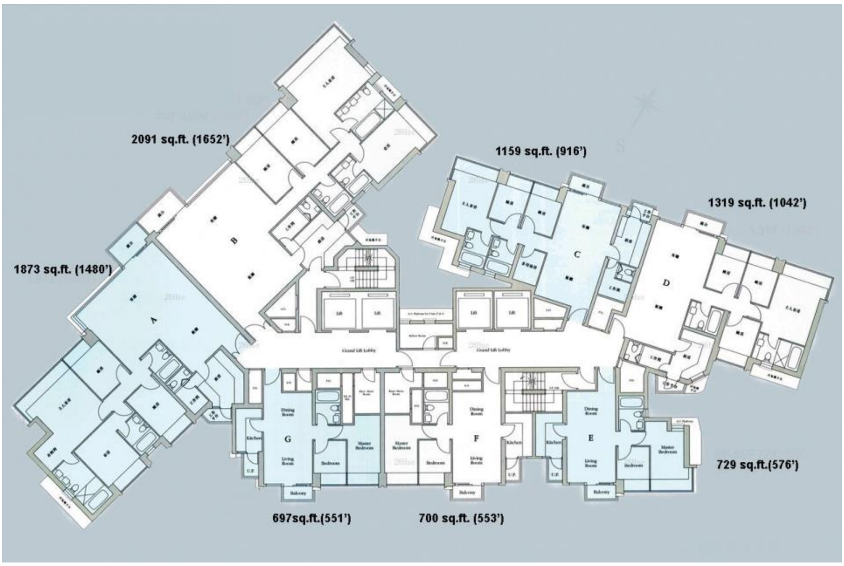
2. Background: Hong Kong Private Housing Market

Hong Kong has one of the most active and volatile real estate markets in the world (Bao and Wan 2004). Residential real estate, in particular, has risen by 19 percent since February 2013 (Spencer and White 2015). More than 53% of the population of seven-million individuals live in private permanent housing while the remaining live in public or temporary housing (Hong Kong Housing Authority 2013). The shortage of housing supply has always been chronic; the Hong Kong government has accordingly started developing new towns since the 1950s to accommodate the booming population. Our dataset is based on all the transactions of private residential units in a new town in Hong Kong named Ma On Shan during the period of July to September 2013, where the median per-square-feet price of a property changed by 0.32% month over month (28hse.com).

There are three reasons for our choice of dataset. First, since the data collection period captures the recovery period after the global subprime crisis in 2008 and the effect of the implementation of cooling measures by the Hong Kong government to curb overheating of investment demand in 2011 and 2012 (Collier 2013), the market was rational and effective. Second, by selecting a narrow time frame (July to September 2013) and a localized district (Ma On Shan), the effects of time-variant (e.g. changes in interest rate, inflation and growth in household income variables) and geographical variables (e.g. proximity to the central business district (CBD)) are minimized (Mok et al. 1995). Hence, the analysis is uncluttered by arbitrary quantifications of macro-level attributes, such as political and geographical differences. Third, the private housing units in Ma On Shan were built in the last twenty years and exhibit varying degrees of structural and auxiliary heterogeneities. Figure 1 shows the heterogeneous structure of the housing units in one private residential estate (Lake Silver) in Ma On Shan. Each floor comprises seven units of various saleable areas (551 to 1652 square feet) and different attributes (e.g. bedroom, storeroom, bathroom, balcony). Figure 2 shows the auxiliary features (e.g. clubhouse, swimming pool, barbecue site, and function halls) that may impact property value and the neighboring housing estates. The unique design and structure of this estate are representative of some of the latest private property developments in Hong Kong and allow us to test our refined model of property valuation with attribute differences in mind.

Previous work in the literature on modeling the Hong Kong real estate market has generally focused on two tracks. One track treats housing as a composite commodity with a vector of complex and heterogeneous features that contribute to the market value (Lancaster 1966; Rosen 1974). The hedonic models study the extent to which features of housing units (e.g. size, age, school district and access to shopping mall) influence property transaction prices. For instance, Mok et al. (1995) apply a hedonic price model to transaction data in Hong Kong and find that a good view of the harbor, large estate size, high floor level, and accessibility to recreational facilities are more desirable housing attributes. This is in line with the analyses done by Choy et al. (2007) and Tse and Love (2000). The former identify the positive effects of property size and floor level on property value; the latter conclude that property values are higher for estate-type housing properties and lower for "dwelling units with a cemetery view". Nevertheless, the hedonic price models still have some issues open for discussion such as the choice of function form, identification of hedonic price functions and data access to the dwelling-level information (Epple, 1987; Ekeland et al. 2004; and Chen and Hao 2010). In real life, such econometric analysis is technically difficult and requires very detailed data for computation. In our present extended framework, we plan to fill this gap by focusing on the list of attributes of a property that are different from its neighbors, instead of an exhaustive list of attributes.

Figure 1 Example of Heterogeneous Structure of Housing Units on A Single Floor in A Private Residential Estate in Hong Kong[1]



[1] The floor plan of Tower One of Lake Silver, a recent large-scale property development in Ma On Shan, is typical of the heterogeneous unit structure across the same floor to cater to a diversity of buyers.

Source: http://www.grandwaterfront.net/thread-37498-1-1.html

Figure 2 Example of Auxiliary Features of A Private Residential Estate in Hong Kong[2]

Another track is drawing reference from the transaction prices in the neighborhood in the recent past when assessing the value of a given property. The emergence of spatial correlation in the error term causes biased parameter estimates, misleading significance levels, and insufficient estimates of the dependent variable (Cliff and Ord 1973; Ripley 1981; Upton and Fingleton 1985; Anselin and Griffith 1988; Odland 1988). According to Can (1990), the spatial correlation in the error term disappears after the spatial variables are introduced. In the literature, the spatial dependence model conventionally finds applications in environmental studies (Ault, Zhong, and Coyle 2005; Cottenie 2005), agricultural field experiments (Besag and Higdon 1999), the social sciences (Xu and Kennedy 2015) and economics (Fujita 2010). Hence, in real estate pricing estimations, the nearest neighbor approach has been introduced in hedonic pricing model frameworks (Can 1990) in order to accommodate the spatial autocorrelation in the error term. As mentioned in Elhorst (2010), a major weakness of spatial dependence modeling is the issue of the spatial weights matrix specification where the matrix is required to be specified in advance. In such a connection, one of the weighting function is applied to gauge the relative importance of its neighbors based on the inverse distance or the

---

[2] The Lake Silver clubhouse features the first private club in Hong Kong with islands on the train station rooftop with director's houses and swimming pools. Other amenities include banquet halls, indoor heated swimming pool, fitness room, outdoor children park, garden barbecue, indoor multipurpose stadium and so on. This auxiliary attribute is a major selling point of the housing estate and a feature that impacts the valuation difference from the neighboring properties.
Source:
http://esfphoto.midland.com.hk/img_wm.php?wm=mr&src=http://img.midland.com.hk/photo_db/outlook/estate/EE000007041_13.jpg

inverse distance squared. The adoption of the nearest neighbor approach is by and large a common practice of lenders and realtors when making decision or giving advice.

Although the application of the nearest neighbor model has made impressive contributions to property valuation in the last two decades, there are still some local conditions that cast some doubt over the prediction accuracy of this model. The validity of the nearest neighbor model is grounded on the assumption that houses in the vicinity share common locational factors and similar structural attributes. However, such a distance dependent weighting function may not be appropriate for measuring the relative importance of neighboring real estates in Hong Kong when the subject and the reference properties are often situated in the same building or neighboring buildings but share different floor plans and own different attributes. Instead of adopting a distance dependent weighting function, we employed a linear-in-means approach proposed by Manski and Manski (1993) and often used to model social interactions or peer effects (Graham and Hahn 2005; Pinkham and Imbens 2011). A linear-in-means approach assumes that the value of a given property can be estimated by the average of the most recent transaction prices of its neighboring properties. Instead of having a complete list of attributes, our modified model integrates linear-in-means value estimation with attribute differences. Due to the heterogeneity of the property market in Hong Kong or its counterparts, we believe that an adjustment of the attribute differences between the given real estate and its neighbors to the value estimated by the linear-in-means complements the deficiency in valuation with the sole use of the nearest neighbor approach. As a special type of spatial dependence model, our approach improves on the conventional model by concentrating on the differences between the attributes of a real estate and its neighbors, thus avoiding some of its shortcomings.

3. Conventional Spatial Dependence Models and Nearest Neighbor Model with Attribute Differences for Residential Property Valuation

This section begins with an introduction of a conventional spatial dependence model, a special type of such a model and finally our nearest neighbor model for residential property valuation for a local region. The conventional spatial dependence model estimates the value of a property by a fraction (ρ) of an average of the transactions of its nearest neighbors, defined by a spatial weights matrix (W), and this estimate is further adjusted by using a linear combination of the attribute values of the property (Elhorst 2010). Instead of using a fraction, we propose to model the value of a property by the average of the transactions of its nearest neighbors, that is $\rho = 1$, with adjustment of the values of the attribute differences between the property and its neighbors.

Throughout this paper, we use the following set of notations. Let m and n be the number of distinct attributes and the number of observed transactions respectively.

Let $P = \begin{pmatrix} P_1 \\ \vdots \\ P_n \end{pmatrix}$ be the $n \times 1$ vector of observed transactions, where P_i $(i = 1,2,\dots,n)$ denotes the log-transaction price per square foot of i^{th} real estate. Since all the properties are distinct, we assume that $P_i \neq P_j$ for all $i \neq j$. Let $F = \begin{pmatrix} F_1' \\ \vdots \\ F_n' \end{pmatrix}$ be the $n \times m$ matrix of attribute selection, where $F_i' = (f_{i1} \cdots f_{im})$ is the attribute value of i^{th} real estate. Let $W = \begin{pmatrix} w_1' \\ \vdots \\ w_n' \end{pmatrix}$ denote the spatial weights matrix, where w_i' defines a tuple of weights to represent the neighborhoods of the i^{th} real estate. For example, $w_4' = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \end{pmatrix}$ implies that the fourth property in the sample has three neighbors. Since the structure of spatial dependence is formulated in the $n \times n$ weight matrix W, each element in W is carefully chosen, such that every property is assigned its nearest neighbors. Equal weights are given to all neighbors, and the matrix is row-normalized. As neighborhood is a bi-directional relation; that is, i is a neighbor of j when j is a neighbor of i, the weight matrix W is symmetric in terms of the positions of its non-zero entries.

Section 3.1 describes various spatial dependence models including our proposed model. Section 3.2 outlines the prior specifications and our Gibbs sampler procedures for various models. Section 3.3 demonstrates the application of the models to property valuation.

3.1 Spatial Dependence Models

A conventional spatial dependence model considers P_i as the sum of three components: (i) a fraction of the average of neighborhood transactions of the i^{th} real estate, (ii) a linear combination of the attribute values of the i^{th} real estate, and (iii) a random error term. Formally, P_i can be written in the form:

$$P_i = \rho w_i' P + F_i' X + \varepsilon_i, \quad i = 1,2,\dots,n, \tag{1}$$

where $\varepsilon_i \sim n(0, \psi_{\varepsilon i})$ and $0 < \rho < 1$.

We now explain the components in (1) in detail. The quantity $w_i' P$ is the average of log-transaction prices per square foot of the neighborhoods of i^{th} real estate. This component captures a portion of the value of i^{th} real estate that is homogeneous to its neighborhoods. It is related to the prevailing market environment and common demographic factors of the district, such as current real interest rates, money supply, access to mass transit railway, types of schools in the district, and access to large shopping malls and recreational facilities. Since $X = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}$ denotes the vector of the attribute values, with X_i representing the value of the i^{th} attribute, for $i = 1, 2, \dots, m$, $F_i'X$ represents another portion of the value of i^{th} real estate that is a linear combination of values of its own set of attributes. Finally, the random error term $\varepsilon_i, i = 1,2, \dots, n,$ denotes any unobserved effects beyond the model specification. This could be, for instance, the bargaining power of the buyer and seller, difference in interior design and any unobserved conditions of the real estates (Bailey 1966).

In matrix form, (1) can be collectively represented as the following matrix equation:

$$P = \rho WP + FX + \varepsilon, \tag{2}$$

where ε is a concatenated vector of all the errors, $\varepsilon_i, i = 1,2, \dots n$. When the model is adequate, ε_is are uncorrelated and the error ε is normally distributed with mean 0 and diagonal covariance matrix $\Sigma_\varepsilon = diag(\psi_{\varepsilon 1}, \dots, \psi_{\varepsilon n})$. The case for $\rho = 0$ corresponds to a multiple regression model of P on F. From this perspective, a conventional spatial dependence model can be considered as an extension to the multiple regression model. Instead of assuming that the price of a real estate is solely driven by a set of features, Equation (2) recognizes that the transactions of the neighborhoods should reflect the expectation of future changes in value due to the factors that affect the district and the entire region. The subsequent adjustment of the initial estimated value of the real estate due to attribute values is assumed to be linear. One may argue that such a linearity assumption is rather restrictive. It is possible that the effects due to attributes can be non-linear (Choy et al. 2007). If the precise functional form is known, an iterative approach can be considered to find parameter estimates. In the following, we will demonstrate that our proposed model with adjustment of feature differences is still applicable when the functional form is unknown.

The second model that we consider, the spatial dependence model with adjustment of feature differences, is to replace the feature vector F_i of the i^{th} real estate in (1) by the feature difference vector $F_i - F'w_i$ between the i^{th} real estate and its neighbors. As the condition $0 < \rho < 1$ is maintained, this is an attempt to keep our proposed model within the context of a conventional family of models. Hence, P_i can be written as:

$$P_i = \rho w_i' P + (F_i - F'w_i)'X + \varepsilon_i, \;\; i = 1,2, \dots, n, \tag{3}$$

where $\varepsilon_i \sim n(0, \psi_{\varepsilon i})$ and $0 < \rho < 1$.

As opposed to (1), the component $(F_i - F'w_i)'X$ in (3) is a portion of the value of the i^{th} real estate that is heterogeneous to its neighbors. This corresponds to the factors such as the existence of a balcony, difference in flat and room sizes, number of bathrooms and facing direction. Here, we assume that potential buyers are often willing to pay extra for an additional attribute (Bao and Wan 2004; Choy et al. 2007), thus suggesting an unobserved value associated with each attribute. Other components of (3) are similar in meaning to that of (1).

Collectively, (3) can be represented by the following matrix equation:

$$P = \rho WP + (I - W)FX + \varepsilon, \tag{4}$$

where ε is a concatenated vector of all the errors, and $\varepsilon_i, i = 1,2, \dots n$, which is normally distributed with mean 0 and diagonal covariance matrix $\Sigma_\varepsilon = diag(\psi_{\varepsilon 1}, \dots, \psi_{\varepsilon n})$.

When $\rho = 1$, the second model becomes our proposed nearest neighbor model with adjustment of feature differences. First, as discussed in the last paragraph, the use of the average of transactions of the nearest neighbors of an apartment as an initial estimate and a subsequent fine-tuned adjustment of this initial estimate with the value of feature differences between the apartment and its nearest neighbors to arrive at a more precise estimation of the value of the apartment, is actually more intuitively appealing. We argue that the natural cause of value differences between an apartment and its nearest neighbors is due to the feature differences rather than the features of the apartment itself. Secondly, the restriction of $\rho < 1$ may not be empirically justified by the local real estate market. The modeling of our Ma On Shan data by the spatial dependence model with adjustment of feature differences suggests that the case of $\rho = 1$ cannot be ruled out. Thirdly, the application of our proposed model does not require a complete list of attributes to compute a value estimate of an apartment. This is definitely an advantage over the conventional model.

Another advantage of the use of a feature difference vector is that the proposed model of adjustment of feature differences is still applicable when the relationship between p_i and F_i is non-linear. Suppose that (3) becomes

$$P_i = \rho w_i' P + (G(F_i) - G' w_i)' Y + \varepsilon_i, \;\; i = 1,2, \dots, n, \tag{5}$$

where $G(F_i) = \begin{bmatrix} G_1(F_i) \\ \vdots \\ G_h(F_i) \end{bmatrix}$ is a $h \times 1$ vector of features, $G = \begin{bmatrix} G(F_1)' \\ G(F_2)' \\ \vdots \\ G(F_n)' \end{bmatrix}$ and Y is a $h \times 1$ vector of unknown parameters. Here, $G(\cdot)$ is a differentiable function of F_i of an unknown form. The j^{th} row of $G(F_i) - G' w_i$ can be written as:

$$G_j(F_i) - \textstyle\sum_{k=1}^{n} G_j(F_k) w_{ik} = \sum_{k=1}^{n} w_{ik} \left(G_j(F_i) - G_j(F_k) \right) \tag{6}$$

After applying the Taylor series on $G_j(F_k)$, (6) becomes

$$G_j(F_i) - \textstyle\sum_{k=1}^{n} G_j(F_k) w_{ik} = \sum_{k=1}^{n} w_{ik} \frac{\partial G_j(F_i)'}{\partial F} (F_i - F_k) \tag{7}$$

$$DG(F_i) = \begin{bmatrix} \frac{\partial G_1(F_i)'}{\partial F} \\ \frac{\partial G_2(F_i)'}{\partial F} \\ \vdots \\ \frac{\partial G_h(F_i)'}{\partial F} \end{bmatrix}$$ is defined as the differential of $G(\cdot)$ with respect to F. (5)

can then be written in the following matrix form.

$$P_i = \rho w_i' P + (F_i - F' w_i)' X + \varepsilon_i, \quad i = 1,2,\ldots,n, \tag{8}$$

where $X = DG(F_i)'Y$. As can be seen, (8) is the same as (3). This confirms that our proposed model with adjustment of feature differences can be used to value a given real estate when an unknown functional relationship between its value and features exists.

Furthermore, as will be discussed in Section 5.3, our proposed model provides a thorough account of the pricing dynamics for both in-sample and out-of-sample predictions.

3.2 Prior Specification and Computation Procedure

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

$$X \sim n(\mu_0, \Sigma_0) \tag{9}$$
$$\psi_{\varepsilon i} \sim IG(\alpha_{0\varepsilon i}, \gamma_{0\varepsilon i}), i = 1, \ldots, n \tag{10}$$
$$\rho|\lambda \sim \text{truncated } n\left(0, \frac{1}{\lambda}\right) \text{ on } (0,1), \lambda \sim \Gamma(u, w) \tag{11}$$

where $\mu_0, \Sigma_0, u, w, \alpha_{0\varepsilon 1}, \ldots, \alpha_{0\varepsilon n}, \gamma_{0\varepsilon 1}, \ldots, \gamma_{0\varepsilon n}$, are hyper-parameters, IG is the inverted gamma distribution and $\Gamma(u, w)$ is the gamma distribution with parameters u and w, respectively.

Given the prior distributions in (9), (10) and (11), our model specification is complete. With these conjugate priors, the full conditional distribution for all parameters and attribute values $[X, \rho, \lambda, \Sigma_\varepsilon | P]$ is of standard form. The Gibbs sampler (Casella and George 1992) is used to sample from them. Sampling from the joint posterior distribution $[X, \rho, \lambda, \Sigma_\varepsilon | P]$ can be obtained by sequentially sampling from each of the following posterior distributions. They are written in accordance with different model specifications as follows.
For the conventional spatial dependence model,

$$X | P, \rho, \Sigma_\varepsilon \sim n(\mu_X, \Sigma_X) \tag{12a}$$

$$\psi_{\varepsilon i}\left|P,X,\Sigma_\varepsilon^{-\varepsilon i},\rho\sim IG\left(\frac{1}{2}+\alpha_{0\varepsilon i},\gamma_{0\varepsilon i}+\frac{1}{2}u_i^2\right),i=1,\dots,n\right. \tag{13a}$$

$$\rho|\lambda,P,X,\Sigma_\varepsilon\sim\text{Truncated } n\left(\frac{P'W'\Sigma_\varepsilon^{-1}(P-FX)}{\lambda+P'W'\Sigma_\varepsilon^{-1}WP},\frac{1}{\lambda+P'W'\Sigma_\varepsilon^{-1}WP}\right)\text{ on }(0,1) \tag{14a}$$

$$\lambda|\rho,P,X,\Sigma_\varepsilon\sim\Gamma\left(u+\frac{1}{2},\frac{1}{2}\rho^2+w\right) \tag{15a}$$

where

$$\mu_X=\Sigma_X(\Sigma_0^{-1}\mu_0+F'\Sigma_\varepsilon^{-1}(I-\rho W)P) \tag{16}$$

$$\Sigma_X^{-1}=\Sigma_0^{-1}+F'\Sigma_\varepsilon^{-1}F \tag{17}$$

$$(I-\rho W)P-FX=\begin{pmatrix}u_1\\ \vdots\\ u_n\end{pmatrix} \tag{18}$$

For a spatial dependence model with adjustment of feature differences, the conditional posterior distributions are the same as those of the conventional model except that $(I-W)F$ replaces F in (10a), (12), (13) and (14) respectively. They are given as follows:

$$X|P,\rho,\Sigma_\varepsilon\sim n(\mu_X,\Sigma_X) \tag{12b}$$

$$\psi_{\varepsilon i}\left|P,X,\Sigma_\varepsilon^{-\varepsilon i},\rho\sim IG\left(\frac{1}{2}+\alpha_{0\varepsilon i},\gamma_{0\varepsilon i}+\frac{1}{2}u_i^2\right),i=1,\dots,n\right. \tag{13b}$$

$$\rho|\lambda,P,X,\Sigma_\varepsilon\sim\text{Truncated } n\left(\frac{P'W'\Sigma_\varepsilon^{-1}(P-(I-W)FX)}{\lambda+P'W'\Sigma_\varepsilon^{-1}WP},\frac{1}{\lambda+P'W'\Sigma_\varepsilon^{-1}WP}\right)\text{ on }(0,1) \tag{14b}$$

$$\lambda|\rho,P,X,\Sigma_\varepsilon\sim\Gamma\left(u+\frac{1}{2},\frac{1}{2}\rho^2+w\right) \tag{15b}$$

where

$$\mu_X=\Sigma_X(\Sigma_0^{-1}\mu_0+F'(I-W)'\Sigma_\varepsilon^{-1}(I-\rho W)P) \tag{19}$$

$$\Sigma_X^{-1}=\Sigma_0^{-1}+F'(I-W)'\Sigma_\varepsilon^{-1}(I-W)F \tag{20}$$

$$(I-\rho W)P-(I-W)FX=\begin{pmatrix}u_1\\ \vdots\\ u_n\end{pmatrix} \tag{21}$$

For our proposed nearest neighbor model with adjustment of feature differences, the conditional posterior distributions are specified as follows:

$$X|P,\Sigma_\varepsilon\sim n(\mu_X,\Sigma_X) \tag{12c}$$

$$\psi_{\varepsilon i}\left|P,X,\Sigma_\varepsilon^{-\varepsilon i}\sim IG\left(\frac{1}{2}+\alpha_{0\varepsilon i},\gamma_{0\varepsilon i}+\frac{1}{2}u_i^2\right),i=1,\dots,n\right. \tag{13c}$$

where

$$\mu_X=\Sigma_X(\Sigma_0^{-1}\mu_0+F'(I-W)'\Sigma_\varepsilon^{-1}(I-W)P) \tag{22}$$

$$\Sigma_X^{-1}=\Sigma_0^{-1}+F'(I-W)'\Sigma_\varepsilon^{-1}(I-W)F \tag{23}$$

$$(I-W)P-(I-W)FX=\begin{pmatrix}u_1\\ \vdots\\ u_n\end{pmatrix} \tag{24}$$

Readers may refer to Casella and George (1992), Choi et al. (2009) and Gelman et al. (2003) for details.

3.3 Model-Based Property Valuation

In each step of the Gibbs sampler, a random number of each parameter is drawn from its conditional posterior distribution given the current value of the other parameters. After a sufficiently large number of Markov chain Monte Carlo (MCMC) iterations, the process is equivalent to sampling from their joint posterior distribution $[X, \rho, \Sigma_\varepsilon, \lambda | P]$. Inference on parameters and the attribute values can be made through their joint posterior distributions. For example, the parameters of the model and the attribute values can be estimated by their posterior means or medians. Their confidence intervals can also be similarly constructed.

The value prediction of an out-of-the-sample real estate is obtained by first identifying its neighbors in the data (recorded transactions), that is, by constructing a vector of weights w, and, secondly, using one of the following model specifications.

For the conventional spatial dependence model:

$$\hat{p} = \hat{\rho} w' P + f' \hat{X}; \tag{25}$$

For spatial dependence model with adjustment of feature differences:

$$\hat{p} = \hat{\rho} w' P + (f - F' w)' \hat{X}; \tag{26}$$

For nearest neighbor model with adjustment of feature differences:

$$\hat{p} = w' P + (f - F' w)' \hat{X}; \tag{27}$$

where $\hat{p}$ is the predicted log-price per square foot of the real estate, w is the vector of weights of the neighbors of the real estate, f is the vector of attributes of the real estate, $f - F'w$ is the vector of attribute differences of the real estate and its neighbors, $\hat{\rho}$ and $\hat{X}$ are posterior means that result from the Gibbs sampler. The value of the real estate is then estimated as the product of $\exp(\hat{p})$ and the gross area of the real estate in square feet.

4. Data Overview

Our dataset contains all of the recorded private residential transactions in a new town area, Ma On Shan, located in the north east part of Hong Kong (see Figure 3).

Ma On Shan has a dense population of approximately 215,300, which is increasing steadily at 1.55% annually (Thomas Brinkoff 2015). Ma On Shan mainly comprises residential areas with over 87% of the population living in private permanent housing and 11% in public housing. There are over 81% of the domestic households who own the quarters that they occupy (The Government of the Hong Kong Special Administrative Region Census and Statistics Department 2012). As in many new towns in Hong Kong, private residential estates are made up of units of different sizes, floor plans and orientations. Figure 1 shows the floor plan from a building in a newly built private residential estate in Ma On Shan. The unit size varies between 551 to 1652 square feet. The units exhibit different attributes in terms of number of bedrooms and bathrooms; the availability of balcony and storeroom, and the facing direction. Figure 2 shows the range of auxiliary features of this estate of which the design is typical of some recently built large-scale private residential estates in Hong Kong. The auxiliary features include a clubhouse, swimming pool, indoor gymnasium, fitness room, snooker room, theatre room, bowling alleys, and function rooms. Other features not shown here include shopping mall, accessibility to public transportation and an international kindergarten in the shopping mall. This unique heterogeneous structure of the buildings and the diversity of auxiliary attributes allow us to model the value of a unit by estate, block, unit, floor specifications and attributes (including structural and auxiliary).

Figure 3 A Map of Hong Kong

We obtained our dataset from a few sources: information from real estate agents, property loan officers and several websites[3] including Centadata, a major property transaction database which contains all of the transactions in Ma On

[3] Our dataset is based on the information from real estate agents, loan property officers and the following websites.

- http://www.citypopulation.de/Hongkong.html?cityid=10602.
- http://www.census2011.gov.hk/en/district-profiles/sha-tin.html.
- http://www1.centadata.cm/cci/cci_e.html

Shan recorded by the Hong Kong Government Land Registry. Figures 4 and 5 show the average prices per square foot and the number of transactions of 32 residential estates in Ma On Shan during the data collection period respectively.

Figure 4 Average Price Per Square Foot for Most Ma On Shan Estates in 2013

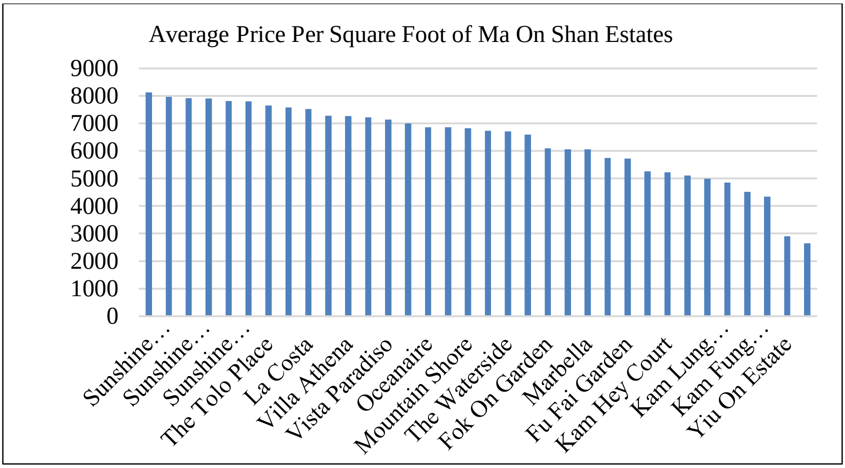
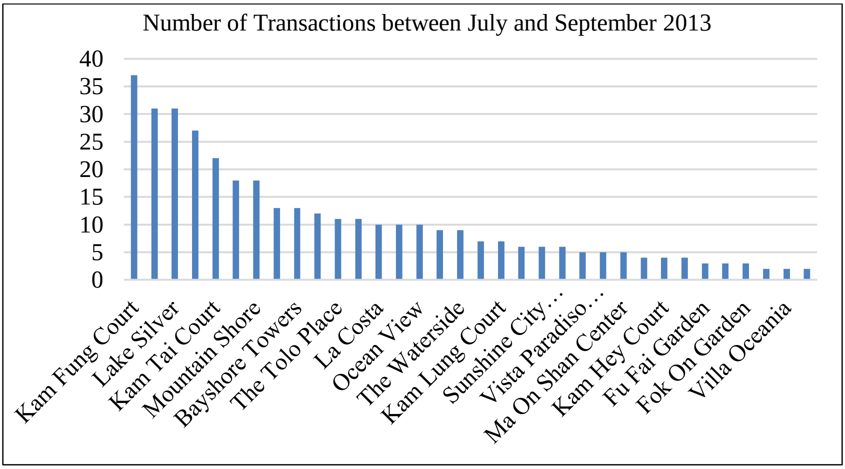


Figure 5 Number of Transactions of Most Estates in Ma On Shan between July and September 2003



A total of 365 registered sale transactions were collected between July and September 2013. Table 1 shows the list of attributes which were obtained from the Hong Kong government offices and census data.

Table 1 Attributes for Property in Ma On Shan

Estate	Name of the Estate Where the Property Is Located
Tower	Identification of the building within the area of an Estate
Floor	The level in a Tower, in which the property is situated
Flat	Identification of the property unit on a Floor
Net Floor Area	Property livable area (in square feet)
Gross Floor Area	Property saleable area (in square feet)
Price per square foot	Transaction price divided by the Gross Floor Area
Date	Date of the transaction
Direction	Direction that property faces (in one of the eight possible directions of north, east, south, west, north east, south east, north west, and south west)
Year Built	The year in which the Tower or the Estate was built
Facilities	Types of facilities provided to residents in an Estate, including bowling, snooker, squash, table tennis, gathering room, outdoor swimming pool, sauna room, golf training course, games room, fitness room, dance room, children's playground, basketball court, gardens, jacuzzi, tennis court, indoor swimming pool
Proximity to Train Station	Measure of neighborhood of property to the popular Mass Transit Railway station (within 15 minutes of walking distance)
Duplex Unit	Yes if the property has a second floor, and No otherwise
Balcony	Yes if there is a balcony in the property, and No otherwise
Ceiling Height	Normal ceiling height of properties in Hong Kong is 8 feet, but some newer properties have ceiling height of 9 to 10 feet.
Bed Room	Number of bed rooms in the property

(Continued…)

(Table 1 Continued)

Estate	Name of the Estate Where the Property Is Located
Bath Room	Number of bath rooms in the property
Multi-Purpose Room	Yes if the property has a multi-purpose room and No otherwise. In some cases, the multi-purpose room is the helper's room.
No. of Units in the Estate	The total number of units in the same estate as the property is located.
Median Age in the Estate	The median age (years) of the population living in the Estate.
Median Income in the Estate	The median income from the main employment of working population living in the Estate.
Usual Spoken Languages by the Residents in the Estate	Usual language spoken by population aged 5 or older in the Estate (for example, Cantonese, Putonghua, English or other languages)
Percentage of Working Population in the Estate	The percentage of population in the Estate, with a fixed place of work in the same area as their residence.
Breakdown Percentages of Employment Types in the Estate	The percentage of employees, employers, self-employed and unpaid family workers in the Estate.
Average household size in the Estate	The average number of persons living together in each of the flats in the Estate
Median Age of Main Family Supporter in the Estate	The median age of the head of households in the Estate

Having compiled a list of attributes for a property, we identify the neighborhood of a property by those in the data that are either situated in the same building or one of the neighboring buildings. Since the local real estate market has sufficient liquidity, properties that satisfy the neighboring condition can always be found. For the purpose of model validation, 40 most recent transactions out of 365 recorded transactions are segregated. Only 325 transactions are used to train our model.

5. Empirical Results

In this section, we calibrate our model to the transaction data described in Section 4. Prior to facilitating a comparison between the conventional spatial dependence model and our proposed nearest neighbor model with adjustment of feature differences, it would be beneficial if our proposed valuation method can be kept within the context of the conventional model. By using MCMC sampling, we first obtain 1,000 samples from the posterior distribution of the model parameters and attribute values based on the spatial dependence model with adjustment of feature differences. We discard the first 500 as "burn-in" samples (Gelman et al. 2003), and retain the remaining 500. The posterior mean and the standard deviation of ρ are found to be 0.998076 and 0.001515 respectively, which results in a 95% confidence interval for ρ as $(0.995046, 1.001106)$. The fact that $\rho = 1$ cannot be ruled out provides further evidence on the applicability of our proposed model to the current data.

In Section 5.1, we summarize the posterior distribution of the model parameters and select attribute values of both the conventional spatial dependence and the nearest neighbor models with adjustment of feature differences. The goodness of fit of both models are validated with the in-sample data. In Section 5.2, we compare the out-of-sample valuation performance of all models considered and the multiple regression model (that is, the case when $\rho = 0$). In Section 5.3, the significant attributes are presented to uncover the secondary factors that are important to homebuyers in the district.

5.1 Model Validation

5.1.1 Conventional Spatial Dependence Model

Having obtained the samples from the joint posterior distribution of the model parameters and the attribute values of the model, the posterior means and 95% confidence interval for the parameter ρ and selected components of attribute vector X are presented in Table 2.

After accounting for the neighboring transactions, the significant attributes are found and listed in Table 2. Contrary to conventional wisdom, the proximity of a shopping mall and the presence of a balcony in an apartment have negative effects on the value estimation of the apartment. However, a positive impact is found on the value of an apartment if it has access to public railway transportation and better net to gross area ratio. The positive values of the attributes "Gross Floor Area" and "Floor Level" imply that a higher gross floor area and higher level in an apartment denote a higher value. Some of these findings, however, are different from the same findings when using our proposed nearest neighbor model with adjustment of feature differences. While the attribute values in the conventional model are in an absolute sense and those in the proposed model are in a differential sense, it is hard to conclude that the apparent deviation between the two models could be due to the existence of correlation between the first and second components of this model described in (1) or (2).

Table 2 Posterior Means and 95% Confidence Intervals for Selected Parameters of Conventional Spatial Dependence Model

Parameter	Posterior Mean	Lower Limit of 95% Confidence Interval	Upper Limit of 95% Confidence Interval
ρ	0.740711	0.666533	0.814889
Gross Floor Area	0.1097	0.0415	0.1778
Floor Level	0.0022	0.0010	0.0034
Balcony	-0.2028	-0.3465	-0.0591
Net and Gross Area Ratio	1.5781	0.8658	2.2904
Shopping Mall	-0.1789	-0.3112	-0.0466
MTR	0.2141	0.1022	0.326
Ceiling Height	0.0765	0.0564	0.0966
Number of Bed Rooms	-0.12	-0.1589	-0.0811

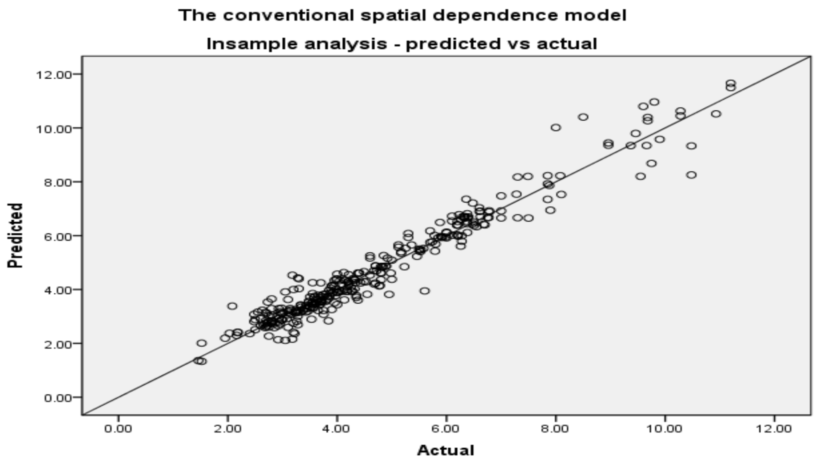
We validate the fit of our model with the in-sample data. Figure 6 presents the observed transaction prices against the fitted valuations for each of the real estates in our dataset that was used to train our model.

Except for some of the apartments with a large gross floor area, the majority of points on the scatterplot in Figure 6 lie close to the 45-degree line, which shows that our model adequately describes the transaction data. The apparent deviation of the transaction prices of apartments with a large gross floor area from the model is due to the implementation of demand control by the Hong Kong government[4] (Collier 2013). The government control mainly affects real estates with a large gross floor area, which causes an increase in the variance of the transaction prices on these real estates. The mean square error on the

[4] Colliers International. 2013, *Cooling Measures in the Hong Kong Real Estate Market: An assessment on their effectiveness and market reaction* [Internet]. Available from: <www.colliers.com>.

transaction prices of the in-sample data is 69.71 and the mean absolute deviation is 104.70, which again indicate a good in-sample fit.

Figure 6 Actual versus Fitted In-Sample Transaction Prices



5.1.2 Nearest Neighbor Model with Adjustment of Feature Differences

Having obtained the samples from the joint posterior distribution of the model parameters and the attribute values of the model, the posterior means and 95% confidence interval for selected components of the attribute vector *X* are presented in Table 3.

Table 3 Posterior Means and 95% Confidence Intervals for Selected Parameters of Nearest Neighbor Model with Adjustment of Feature Differences

Parameter	Posterior Mean	Lower Limit of 95% Confidence Interval	Upper Limit of 95% Confidence Interval
Gross Floor Area	-0.228	-0.344	-0.112
Floor Level	0.0018	0.0008	0.0028
Bath Room	0.0687	0.0271	0.1103
South Facing	-0.0481	-0.0954	-0.0008
Balcony	0.4751	0.0113	0.9388

Since the attribute values “Gross Floor Area” and “South Facing” are negative, the unit square foot price of an apartment with a large gross floor area or a south-facing direction is negatively influenced. On the other hand, a positive impact is found on the unit square foot price of apartments with a balcony, an extra bathroom and a higher level in the building. Since the value attached to each attribute could be subject to many changes such as homebuyer preferences and government policies, these findings should not be generalized as universal. These changes are temporal bound and the value of the attributes is time-dependent. It is therefore unwise to generalize the findings from this study to other cases, but the statistical framework is a useful tool for assessing the relative importance of given attributes at the time of assessment.

We validate the fit of our model with the in-sample data. Figure 7 presents the observed transaction price against the fitted valuation for each of the real estates in our dataset that was used to train our model.

Figure 7 Actual versus Fitted In-Sample Transaction Prices

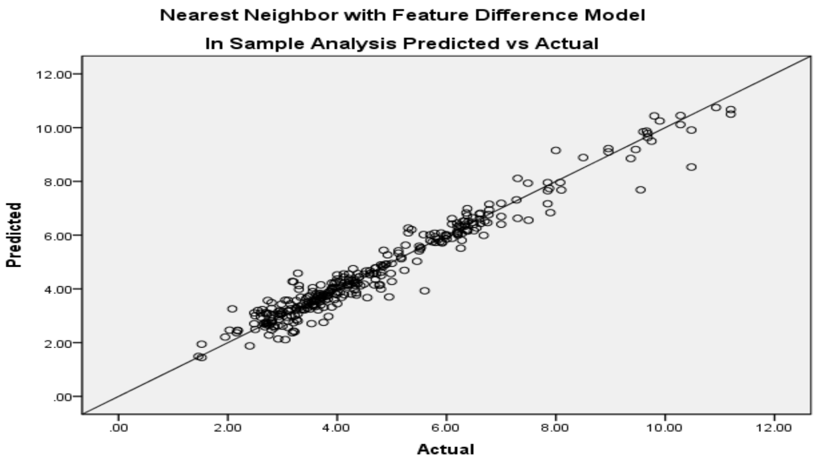
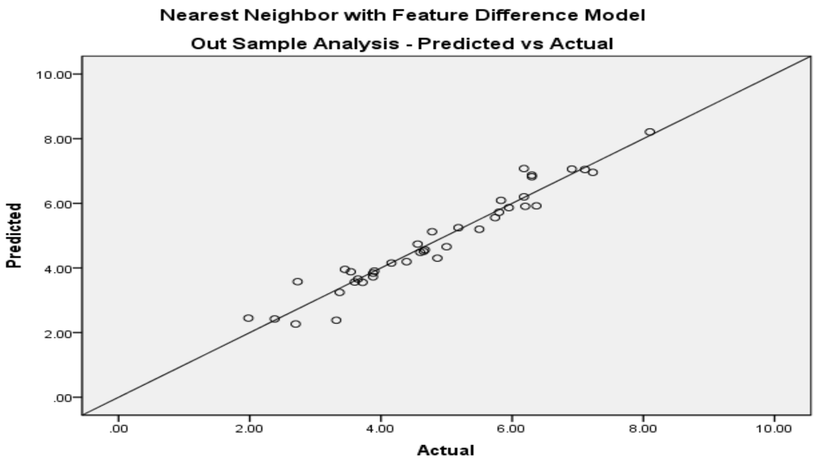


Figure 7 suggests a similar in-sample fit between the two models. However, it is worth mentioning that although the data have a higher variance on the apartments with higher gross saleable area, the current model gives a better fit than that of the conventional model, as shown in the right tail of Figure 7, where the points are closer to the line $x = y$. The mean square error on the transaction prices of the in-sample data is 56.88 and the mean absolute deviation is 94.35, which again indicate a good in-sample fit. Since the mean square error and the mean absolute deviation of the current model are 80% and 90% of that of the conventional model respectively, this shows that the current model has a better in-sample fit to the data than that of the conventional model.

5.2 Out-of-Sample Model Validation

The most recent forty recorded transactions in the dataset were reserved for an out-of-sample validation of our model. Our model was then estimated based on the remaining transactions and used to predict the valuation of the forty reserved apartments. For each of the forty out-of-sample properties, their neighbors in the rest of the dataset and their attribute differences were identified. Then Equation (27) was used to predict its log unit square foot price and the corresponding predicted transaction price of the property was evaluated by the product of the predicted price per square foot and the gross floor area of the property. Figure 8 displays the result of the comparison between the actual and the predicted transaction prices of the forty out-of-sample properties.

Figure 8 Actual versus Predicted Out-of-Sample Transaction Prices



The dotted line in Figure 8 is $y = x$. Points that are closer to the dotted line show the better predictability of our model. As shown in Figure 8, our model performs well in terms of predicting the forty out-of-sample transactions. The performance of our model is further compared with that of three other valuation methods: namely the conventional spatial dependence model discussed in Section 3, spatial dependence model with adjustment of feature differences, and multiple regression model (corresponds to $\rho = 0$ in the conventional model). The average prediction error sum of the squares of the forty out-of-sample transactions with the four models are presented in Table 4.

As seen in Table 4, the average out-of-sample prediction error sum of squares of the nearest neighbor model with attribute differences is the smallest among the four models, hence suggesting that the proposed model is an effective one. Since the 95% confidence interval for ρ includes one, the two models are statistically indifferent. However, the multiple feature regression model is the least effective among the four models irrespective of in-sample or out-of-sample valuations.

Table 4 Average Prediction Error Sum of Squares of Various Models

Dataset	Conventional Spatial Dependence Model ($\hat{\rho} = 0.741$)	Spatial Dependence Model with Feature Differences ($\hat{\rho} = 0.998$)	Nearest Neighbor with Feature Difference s($\rho = 1$)	Regression Model with Multiple Features ($\rho = 0$)
Training	69.711	60.27	56.88	156.12
Validation	5.84	5.14	5.12	13.81
Total	75.55	65.42	62.01	169.93

5.3 Value of Significant Features

Since the transactions of the nearest neighbors of a property have captured the macro-level variables, the residual values are the differences in attributes that distinguish its neighbors from the given property. Hence, the attributes with a significant value should provide invaluable information to homeowners of the district. Table 5 presents the posterior means and their 95% confidence intervals of 10 attributes.

Table 5 Posterior Means and 95% Confidence Intervals for Selected Attribute Values

Attribute	Posterior Mean	Lower Limit of 95% Confidence Interval	Upper Limit of 95% Confidence Interval
Gross Floor Area	-0.228	-0.344	-0.112
Level in the Building	0.0018	0.0008	0.0028
Bath Room	0.0687	0.0271	0.1103
South Facing	-0.0481	-0.0954	-0.0008
Balcony	0.4751	0.0113	0.9388
Net to Gross Area Ratio	0.866	-0.300	2.031
Year Built	0.219	-0.104	0.542
Number of Bed Rooms	-0.019	-0.061	0.023
North Facing	0.015	-0.030	0.060
Ceiling Height	0.016	-0.012	0.044

The five significant attributes identified are the gross floor area, level in the building, balcony, south-facing direction and an additional bathroom. First, at the time the data were collected, a greater gross floor area meant a lower unit square foot price. For example, a 1700 square-foot apartment could be sold for USD836.00 per square foot while a 750 square-foot apartment for USD965.00

per square foot. This could be a repercussion of the implementation of the cooling measures by the government such as the Special Stamp Duty (SSD), a restriction period of selling and the first introduction of the Buyers' Stamp Duty (BSD) in an attempt to curb excessive property speculation and short-term trading activities.[5] Second, our analysis also confirms the general perception that an apartment with a balcony has a better selling price than one without a balcony (Chau 2005). Our model also confirms the general perception that higher floors of an apartment have a higher value. However, contrary to conventional wisdom, a south-facing apartment in Ma On Shan is ranked inferior to those facing other directions for the reason that the units with a sea view are not south facing. Finally, a less mentioned attribute in the literature is "additional bathroom". Our analysis indicates that homeowners prefer more than one bathroom in an apartment.

When compared to the significant attributes identified by the conventional spatial dependence model (see Table 2), those by our model tend to more distinct, less overlapping and easier to access. The attributes identified by the conventional model such as racquetball and table tennis facilities, banquet hall, game room, shopping mall, basketball court, and MTR, are categorically auxiliary attributes which do not help to account for the price difference of units within the same residential estate or residential estates that share similar auxiliary features. Other features, e.g. ceiling height and net to gross area ratio, require information from property developers. In comparison, our proposed model gives significant attributes that are more unique to the subject property and more accessible to homebuyers.

6. Conclusion and Future Research

In this paper, an extended approach for modeling property value in primarily two steps has been presented: (i) the use of a nearest neighbor approach and (ii) adjustment of attribute differences. We have calibrated our model on a dataset that includes all the property transactions in a new town in Hong Kong for a 3-month period in 2013. Our modified model has taken into account the macro-level attributes e.g. locational, physical and economic variables, and the micro-level attributes e.g. structural and auxiliary variables. This model overcomes the possible issues of infrequent transactions of similar neighboring housing units and accounts for the inherent minor differences between similar neighbors. An adjustment of attribute differences is shown to improve the nearest neighbor approach and enhance valuation performance. Our study also shows that the average in-sample and out-of-sample prediction error sum of squares of the nearest neighbor model with attribute differences is the smallest among the four

[5] The details of the cooling measures in the Hong Kong real estate market by the Hong Kong government introduced in the period of 2010 to 2012 are found in Collier's report titled *Cooling Measures in the Hong Kong Real Estate Market: An assessment on their effectiveness and market reaction*. Available from: www.colliers.com [March 2013].

models studied. The significant attributes that are important to homeowners of the district can be identified, which serve as a useful reference for investors or potential homeowners of a localized district. Lastly, a point worthy of mention is that the proposed method of valuation is intuitively simple, easy to understand as well as convenient to implement.

Our proposed two-step approach has a few limitations. First, given the model only considers spatial dependence and attribute differences, it is only suitable for data of a short span where fluctuation of prices is not high. Second, the model is also not suitable for a period where the error variance is not constant, for instance, in a period immediately after the implementation of a government intervention policy or natural disaster. The aftermath period of turbulence is likely to increase the error variance, thus upsetting the effectiveness of the model. Finally, the model is only suitable to value a property in a place where free trade is allowed as the market is governed by the preferences of buyers and sellers.

Future research may consider extending our model framework by directly incorporating a temporal effect on the attribute values. As changes in the value of a property can be viewed as a spatial-temporal process (Valentini, Ippoliti and Fontanella, 2013), our study only considers the spatial nature of a property. The attribute values can in fact be modeled by a Markov process to measure the dynamic changes in prices in the district over time. Combining the two analyses together can frame and exploit both the spatial and the temporal structures of the observed process.

To conclude, we believe that this paper provides a formal exposition of an intuitive and refined two-step model for objective valuation in a local region of the Hong Kong property market and intended to benefit home-buying stakeholders and enhance valuation performance.

References

28Hse.com (2015), *Ma On Shan Property Indices* [Internet]. Available from: <http://data.28hse.com/en/datarecord101.html>

Anselin, L. and Griffith, D.A. (1988), Do Spatial Effects Really Matter in Regression Analysis?, *Papers in Regional Science*, 65, 1, 11-34.

Ault, A., Zhong, X., Coyle, E.J. (2005), K-Nearest-Neighbor Analysis of Received Signal Strength Distance Estimation Across Environments, Proceedings of the First Workshop on Wireless Network Measurements.

Bailey, M. J. (1966). Effects of race and of other demographic factors on the values of single-family homes. *Land Economics*, 42(2), 215-220.

Bao, H.X. and Wan, A.T. (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.

Bartlett, M.S. (1978), Nearest Neighbour Models in The Analysis of Field Experiments. *Journal of the Royal Statistical Society. Series A. Statistics in Society*. 147.

Besag, J., & Higdon, D. (1999). Bayesian analysis of agricultural field experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 61(4), 691-746.

Brinkoff, T. (2015), *City Population* [Internet]. Available from: <http://www.citypopulation.de>

Can, A. (1992), Specification and Estimation of Hedonic Housing Price Models, *Regional Science and Urban Economics,* 22, 3, 453-474.

Can, A. (1990), The Measurement of Neighborhood Dynamics in Urban House Prices, *Economic Geography*, 254-272.

Caplin, A., Chopra, S., Leahy, J.V., LeCun, Y. and Thampy, T. (2008), Machine Learning and The Spatial Structure of House Prices and Housing Returns, *Available at SSRN 1316046*.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46(3), 167-174.

Centadata Company Limited. (2015), *Centadata - New Territories East - Ma On Shan*. [Internet]. Available from: www1.centadata.com.

Chau, K.W. (2005), Real Estate Price Indices in Hong Kong, *Journal of Real Estate Literature,* 13, 3, 337.

Chen, J. and Hao, Q. (2010), Submarket, Heterogeneity and Hedonic Prediction Accuracy of Real Estate Prices: Evidence from Shanghai, *International Real Estate Review,* 13, 2, 190-217.

Choi, J., Hui, S. K., & Bell, D. R. (2008). Bayesian Spatio-Temporal Analysis of Imitation Behavior Across New Buyers at an Online Grocery Retailer. *Journal of Marketing Research*, 45.

Choy, L.H., Mak, S.W. and Ho, W.K. (2007), Modeling Hong Kong Real Estate Prices, *Journal of Housing and the Built Environment,* 22, 4, 359-368.

Cliff, A.D. and Ord, J.K. (1973), *Spatial Autocorrelation,* Pion London.

Colliers International. (2013), *Cooling Measures in the Hong Kong Real Estate Market: An assessment on their effectiveness and market reaction* [Internet]. Available from: <www.colliers.com>.

Cottenie, K. (2005), Integrating Environmental and Spatial Processes in Ecological Community Dynamics, *Ecology Letters*, 8, 1175-1182.

Craig, M.R.S. and Hua, M.C. (2011), *Determinants of Property Prices in Hong Kong SAR: Implications for Policy,* International Monetary Fund.

Damodaran, A. (2008), "Introduction to Valuation" in *Handbook of Finance* John Wiley & Sons, Inc..

Damodaran, A. and Damodaran, (2012) A. *Investment Valuation: Tools and Techniques for Determining The Value of Any Asset,* 3; 3ed, Wiley; Wiley, Hoboken, N.J.; Hoboken, N.J.

Ekeland, I., Heckman, J.J. and Nesheim, L.P. (2003), Identification and Estimation of Hedonic Models, *Journal of Political Economy*, 112, S1, Papers in Honor of Sherwin Rosen: A Supplement to Volume 112 (February 2004), pp. S60-S109

Elhorst, J.P. (2010), Applied Spatial Econometrics: Raising the Bar, *Spatial Economic Analysis*, 5, 1, 9-28.

Epple, D. (1987), Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products, *The Journal of Political Economy*, 59-80.

Fujita, M. (2010), The evolution of Spatial Economic: From THÜNEN to the New Economic Geography, *The Japanese Economic Review*, 61, 1.

Ge, J., Runeson, G. and Lam, K. (2003), Forecasting Hong Kong Housing Prices: An Artificial Neural Network Approach, *International Conference on Methodologies in Housing Research, Stockholm, Sweden*.

Gelfand, A.E., Ghosh, S.K., Knight, J.R. and Sirmans, C.F. (1998), Spatio-Temporal Modeling of Residential Sales Data, *Journal of Business & Economic Statistics*, 16, 3, 312-321.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian Data Analysis (Vol. 2).* Boca Raton, FL, USA: Chapman & Hall/CRC.

Goldsmith-Pinkham, P. and Imbens, G.W. (2013), Social Networks and The Identification of Peer Effects, *Journal of Business & Economic Statistics,* 31, 3, 253-264.

Graham, B.S. and Hahn, J. (2005), Identification and Estimation of The Linear-In-Means Model Of Social Interactions, *Economics Letters,* 88, 1, 1-6.

Griffith, D.A. (1992), What is Spatial Autocorrelation? Reflections on The Past 25 Years of Spatial Statistics, *Espace Géographique*, 21, 3, 265-280.

Griffith, D.A. (1987), Spatial Autocorrelation, *A Primer (Washington, DC, Association of American Geographers)*.

Hong Kong Housing Authority Annual Report 2013/14 (2014), https://www.housingauthority.gov.hk/mini-site/haar1314/en/index.html

Hui, E.C. and Wong, J.T. (2009), The Forecasting Capacity of Housing Price Expectations, *International Real Estate Review,* 12, 1, 39-61.

Hui, S.K., Cheung, A. and Pang, J. (2010), A Hierarchical Bayesian Approach for Residential Property Valuation: Application to Hong Kong Housing Market, *International Real Estate Review,* 13, 1, 1-29.

Ioannides, Y.M. (1999), Economic Geography and The Spatial Evolution of Wages in the United States. In L. Anselin and R. Florax (eds), Advances in Spatial Econometrics. Heidelberg: Springer-Verlag.

Isaac, D. and O'Leary, J. (2012), *Property Valuation Principles,* Palgrave Macmillan.

Isakson, H. (1988), Valuation Analysis of Commercial Real Estate Using the Nearest Neighbors Appraisal Technique, *Growth and Change,* 19, 2, 11-24.

Lam K. (2008), *Clubhouse Facilities in Private Residential Development : An Actual Demand or a Symbol of Identity*, University of Hong Kong (Pokfulam Road, Hong Kong). (Dissertation)

Lancaster, K. J. (1966), A new approach to consumer theory. *The Journal of Political Economy*, 132-157.

Manski, C. and Manski (1993), Identification of Endogenous Social Effects: The Reflection Problem, *The Review of Economic Studies,* 60, 3, 531.

Mok, H.M., Chan, P.P. and Cho, Y. (1995), A Hedonic Price Model for Private Properties in Hong Kong, *The Journal of Real Estate Finance and Economics*, 10, 1, 37-48.

Odland, J. (1988). *Spatial Autocorrelation*. Sage, Newbury Park, California

Ripley, B. D. (1981). *Spatial Statistics. 1981*. Wiley, New York.

Rosen, S. (1974), Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *The Journal of Political Economy*, 34-55.

Spence, P. and White, A. (2015). '*Mapped:* Where in The World are House Prices Growing Fastest?' *Spencer and White UK*, 11 June. Available from: < www.Spencer and White.co.uk/finance/property/11667219/Mapped-Where-in-the-world-are-house-prices-growing-fastest.html>.

The Government of the Hong Kong Special Administrative Region Census and Statistics Department. (2012).

Tse, R.Y. and Love, P.E. (2000), Measuring Residential Property Values in Hong Kong, *Property Management*, 18, 5, 366-374.

Tse, Y., Ho, C. and Ganesan, S. (1996), *An Econometric Analysis of House Prices in Hong Kong*, Department of Architecture, University of Hong Kong.

Upton, G. and Fingleton, B. (1985), *Spatial Data Analysis by Example. Volume 1: Point Pattern and Quantitative Data.* John Wiley & Sons Ltd.

Valentini, P., Ippoliti, L. and Fontanella, L. (2013), Modeling US Housing Prices by Spatial Dynamic Structural Equation Models, *The Annals of Applied Statistics*, 7, 2, 763-798.

Xu, Y., Kennedy, E. (2015), An Introduction to Spatial Analysis in Social Science Research, *The Quantitative Methods for Psychology*, 11, 1, 22-31.