Routledge
Taylor & Francis Group

INTERNATIONAL JOURNAL OF STRATEGIC PROPERTY MANAGEMENT
ISSN 1648-715X / eISSN 1648-9179

2017 Volume 21(3): 273–283
doi:10.3846/1648715X.2016.1259190

FORECASTING SPATIAL DYNAMICS OF THE HOUSING MARKET USING SUPPORT VECTOR MACHINE

Jieh-Haur CHEN [a], Chuan Fan ONG [b], Linzi ZHENG [c], Shu-Chien HSU [d,*]

[a] *Institute of Construction Management and Management, National Central University, Taiwan*
[b] *Department of Civil Engineering, Universiti Tunku Abdul Rahman, Malaysia*
[c] *Department of Real Estate and Construction, The University of Hong Kong, Hong Kong, China*
[d] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong, China*

Received 20 November 2015; accepted 1 September 2016

ABSTRACT. This paper adopts a novel approach of Support Vector Machine (SVM) to forecast residential housing prices. As one type of machine learning algorithm, the proposed SVM encompasses a larger set of variables that are recognized as price-influencing and meanwhile enables recognizing the geographical pattern of housing price dynamics. The analytical framework consists of two steps. The first step is to identify the supporting vectors (SVs) to price variances using the stepwise multi-regression approach; and then it is to forecast the housing price variances by employing the SVs identified by the first step as well as other variables postulated by the hedonic price theory, where the housing prices in Taipei City are empirically examined to verify the designed framework. Results computed by nonparametric estimation confirm that the prediction power of using SVM in housing price forecasting is of high accuracy. Further studies are suggested to extract the geographical weights using kernel density estimates to reflect price responses to local quantiles of hedonic attributes.

KEYWORDS: Housing price forecasting; Spatial dynamics; Supporting vector machine; Hedonic appraisal method

1. INTRODUCTION

The residential housing is not only providing living spaces for people in societies, but offers attractive markets for investors to participate, particularly for international metropolises. Facing fluctuating economies, prospective participants of the property market have been keenly aware of the varying appreciation rates (Haughwout *et al.* 2011). Housing prices forecasting hence is one of the most prominent focal subjects in both the market and the academia. From a methodological perspective, existing pricing forecasting models in the housing market domain could be roughly divided into theory-driven models and data-driven models (Larson 2010). Extensive explorations of questions that are associated with forecasting market directions have been strived, including "where is the turning point of housing pricing variations (Bracke 2013)?; how to capture and forecast the peak of housing prices with unique informative variables (Wu *et al.* 2014)?; and how to forecast the market directions with the lowest error (Beracha, Wintoki 2013)?". Advanced techniques for housing prices forecasting are simultaneously enriched, such as the autoregressive (AR) models (Wu, Brynjolfsson 2013), the autoregressive fractionally integrated moving average (ARFIMA) model (Aye *et al.* 2014), the unobserved components (UC) model (Kishor *et al.* 2015), the random acceleration model (Barari *et al.* 2014), the vector auto-regression (VAR) (Aye *et al.* 2015), and the vector error correction (VEC) model (Wheaton *et al.* 2014).

Though these various models employ different information sets and have been tested by advanced techniques, they generally share a common nature in terms of the hedonic appraisal method. The hedonic price model has its foundation in consumer theory and has been the standard tool of housing appraisal approaches. It considers that an urban area can be treated as a single market for housing services (Adair *et al.* 2000). As also argued by

* Corresponding author. E-mail: mark.hsu@polyu.edu.hk

Rosen (1974), "Goods are valued for their utility-bearing attributes", Indeed, housing price largely depends on, if not defined by, various housing characteristics. As long as a hedonic model can be built to encompass all effective variables, including those that can capture the market supply and demand, the housing price can always be predicted accurately enough.

Traditionally, the econometric relationship between the price and the property characteristics is determined via multiple regression analysis (MRA). Unfortunately, due to the impossibility of building a hedonic that can embrace all effective information sets, the traditional MRA-based hedonic has an inherent vulnerability. In addition, it is structurally bound to assuming an a priori functional relationship between sale prices and property attributes. Furthermore, when there are ever-volatile periodical movements involving multi-attributed affecting factors both endogenously and exogenously, the MRA approach can do nothing to capture the information of the price formation process. The dynamic price formation, however, is essential for accurately modeling the complex dynamics of the housing market. Finally, recent studies suggest that the MRA-based hedonic may lead to biased results if real estate prices are spatially correlated (Osland 2010; Shin *et al.* 2007). These specified four failures would result in untenable or imprecise coefficients caused by functional form misspecification, interaction among variables, multicollinearity, and non-linearity problems (Zurada *et al.* 2011). As witnessed by a sight of familiar truths, market directions were often unsatisfactorily forecasted on the one hand, and spatial price dynamics were empirically understudied on the other. To date, both industry and academia call for innovations in forecasting housing price variations with the complexities and dynamics of the housing market as main concerns.

Recently, machine learning algorithms, or computational intelligence approaches, have attracted increasing attention for mass appraisal in stock markets. As a potentially more effective and flexible approach for price forecasting, machine learning algorithms also have been applied in the real estate domain. For examples, González and Formoso (2006) developed mass appraisal models using fuzzy logic; Nghiep and Al (2001) indicate that artificial neural networks (ANN) perform better compared to MRA for large sample sizes, and d'Amato (2007) employs rough set theory as an automated valuation method. Out of excising machine learning technologies, the Supporting Vector Machine (SVM) is of particular usefulness for it can learn the input-out functionality from existing samples, and is capable of mapping the special dynamics into a high dimensional feature space (Cristianini, Shawe-Taylor 2000). Therefore, in this paper, a novel model is built that combines the typical hedonic appraisal method with the support vector machines (SVMs) to generate a mass appraisal model that could rest on theoretical expectation and at the same time circumvent the restrictions of MRA.

Remaining contents are organized as follows. Section 2 of this paper provides a critical review of the literature associated with housing price forecasting; section 3 describes the research strategy. Section 4 presents the empirical strategy of applying SVM to forecast spatial dynamics in housing price of Taipei and Section 5 provides analytical discussions. Section 6 concludes this paper and suggests future explorations.

2. LITERATURE REVIEW

2.1. Cross-sectional differences of housing price

A number of factors have been recognized by researchers as can influence housing prices significantly. From the view of cross-sectional differences in housing price, housing appreciations and variances are widely documented in the previous literature. For example, Mayer (1993) finds that high price houses appreciate at a higher rate in four U.S. cities, but also they suffer more volatile than lower price houses. Smith and Ho (1996) confirm the hypothesis that the monetary shocks widen the price differential between higher and lower priced housing properties, while the real shocks (e.g. variations in real income and employment rate) narrow the price differential. They further point out that the nature of the economic cycles and local market conditions can also affect the differential between lower and higher priced properties. Drake (1993) tries to find determinates of housing prices by performing Johansen cointegration analysis (CA) on housing stock and mortgage interest rates in the British housing market. De Greef and De Haas (2000) apply the CA and VEC to analyze relationships between the mortgage market and housing market in Netherlands.

Regarding the demographical and economic variables, researchers reveal that different variables tend to have different impacts on the dynamic behavior of housing prices and the number

of houses sold in different regions at different periods. For example, Ohtake and Shintani (1996) use a housing price model with demographic factors to analyze how the baby boom and the independence boom of youngers affect Japanese housing price; Holly and Jones (1997) examine extensive long-run panel data from the year 1939 to 1994 for the British market to analyze how real income, demographic factors, and interest rates affect housing stock; and Lamont and Stein (1999) find that housing prices in cities, where a large proportion of homeowners is highly leveraged, are more sensitive to external economic shocks. In De Greef and De Haas (2000)'s study, variables such as real income, population, housing stock, housing mortgage interest rate, and amount of real income expectation are adopted and concluded as affecting housing price variances differently in Netherlands.

2.2. Previous studies on housing price forecasting

Given that housing prices are varying and being influenced by extensive variables, many researchers have attached great attentions on how to forecast housing prices accurately. Earlier studies on housing price formulation and forecasting mainly follow the basic law of demand and supply combined with the adjustment latency, and the demand side is assumed to be the major determinate of housing prices during short periods. Such concentrations on modelling market structures, such as studies undertaken by Whitehead (1974), Hadjimatheou (1976), Mayes (1979) and Hendry (1984), nonetheless turn out to be weak in forecasting accuracy. Time series analysis is very different from these structural models, the basic proposition of which is 'history might matter'. Case and Shiller (1989) adopt a time series analysis method and examine the repeat sales price index in 49 cities of the US to test whether the American residential housing market is consistent with the weak-efficient market hypothesis or not. Muellbauer and Murphy (1997) use this method and find that because of the unstable relationships among households' wealth, households' income, interest rates, and home prices, the housing prices could move dramatically as time changes. Brown *et al.* (1997) believe the ignorance of structural changes in the housing market is the major reason for failure in house price forecasting. Therefore, Brown *et al.* (1997) propose a Time Varying Coefficient (TVC) method with an unstable price data production assumption and get better regression results of varying coefficients compared to three other regression results derived from constant coefficients in models of ARIM, ECM, and VAR. In following with Brown *et al.* (1997)'s arguments about market structure, Ortalo-Magne and Rady (2006) develop a regime-switch model of housing to analyze the pattern of housing price movements. They find that the financial constraints of Youngers are major reasons for the nonlinearity changes of housing prices.

Although previous studies on housing price dynamics are rich, they dedicate disproportionate attention to technology, leading to grave shortcomings in theory analysis and outcomes of which that are often contrary to each other. The inconsistency of results, and thus lower accuracy in price forecasting are challenges in the research focal of housing price forecasting. As a novel technology of learning machine, the Supporting Vector Machine (SVM) has been proved to be able to solve problems of limited sample learning, nonlinear regression, and overcome the "curse of dimensionality", and thus causes emerging attentions of researchers in price forecasting in various markets including the real estate market (e.g., Kazem *et al.* 2013; Wang *et al.* 2014; Zhang *et al.* 2015). The capabilities of SVM also initiates this article to solve the spatial dynamics of housing price in Taiwan.

2.3. Housing price determinants

According to Lancaster (1966)'s consumer theory, goods possess bundles of characteristics and it is those characteristics, not the goods themselves, on which the utility is derived. Utility or preference are assumed to rank bundles of characteristics directly and only rank collections of goods indirectly based on their characteristics. One single good may possess a bundle of characteristics, and those characteristics work together as a group. Purchase of such goods means the acquisition of those characteristics and converting them into utility. This kind of market should be described with a range of qualities or the prices of the characteristics that the goods contain, which we will call the hedonic price. As being put by Rosen (1974), "Goods are valued for their utility-bearing attributes or characteristics", thus the housing price can be considered being determined by collections of hedonic values and different attributes have different implicit prices.

Rosen (1974) proposed market supply and demand equilibrium to describe product attributes. Under the assumption of perfect competition and aiming to maximize consumer utility, Rosen (1974)

analyzed the short- and long-term equilibriums of the heterogeneity goods market and established a sound base for model specification and estimation of the Hedonic Price Model. According to Rosen's theory, econometric methods can be applied to derive the implicit price of product characteristics. The observed product price and the bundles of characteristics of the good define a set of implicit or hedonic prices.

Building on the above-mentioned theories, the most common hedonic variables of the property include the age, type of property, floor area, number of living halls, the number of rooms, the number of garages, and other amenities available within the property. In addition, no two properties share the same space and each property is spatially bound to its locational features within a neighborhood or sub-market, such as differences in transportation, accessibility, public services, aesthetic quality, recreational facilities, etc. The locational attributes generate urban externalities which can be capitalized on the market but sometimes might also impair the property value. Accessibility has been regarded as a major influence in the modeling of residential location. According to Hanson (2004), accessibility refers to the number of opportunities available within a certain distance or travel time. Following Hanson (2004), various measures have been proposed and used in previous studies to improve the precision of measuring accessibility; with more recent studies able to take advantage of GIS. For example, Song and Sohn (2007) propose that accessibility can be measured according to the percentage of specific land use, or measured as the path distances from the housing unit to the Central Business District (CBD), or constructed with a more complex accessibility index that incorporates both distance and size of targeted facilities (Song, Sohn 2007).

In most cases, the CBD is regarded as the center for many activities. Therefore, proximity to the CBD is considered an attractive quality that increases property prices. Nonetheless, transportation infrastructure can enhance the mobility of residents and possibly reduce demand friction around the CBD, so that residential property can generate value from its proximity and superior access to transportation infrastructure (Debrezion *et al.* 2007). The effect of transportation on property value will vary depending on the configuration of transportation modes and networks, such as the availability of highways, light or mass rapid transit, bus transit, etc. By and large, studies concur that the ease of transportation has a positive impact on the property value, though the degree of impact differs across the literature, depending on the stratification level and mode of transportation (see Adair *et al.* 2000; Billings 2011; Hess, Almeida 2007).

Accessibility and size of retail stores are two key traits examined in research on retailing and housing price. For instance, studies indicate that the size of retail stores positively impact residential values (Sirpal 1994), while greater accessibility to retailing has a positive effect on housing price but the effect diminishes if the distance of a store is too close to a house (Des Rosiers *et al.* 1996; Song, Sohn 2007).

Environmental amenities have important aesthetic value and provide opportunities for recreational activities. The amenity values comprise of two components. The first is accessibility to the resource, such as a park or beach, that contributes to recreational and leisure purposes. The second value is attributed to the scenery view of resources (Hamilton, Morgan 2010). In the same vein, Kong *et al.* (2007) show that green space area, and accessibility to plazas and parks positively impact residential property value. On the other hand, cultural amenities such as museums, zoos, theaters, symphonies and dance companies, also positively affect the locational choice of workers (Clark, Kahn 1988).

With regard to public services such as schools, hospitals, fire departments, and police stations, little a priori explanation has been made relating their accessibility to property value. Nevertheless, the property taxes collected in a municipality could translate into public services that should have a positive effect on property value (Kauko 2003). Accessibility of hospital, elementary school, university are usually adopted in hedonic price modeling (Waddell *et al.* 1993; Matthews, Turnbull 2007; Kryvobokov, Wilhelmsson 2007; Keskin 2008), while studies by Gibbons and Machin (2008) and Billings (2015) highlight school quality and lower crime rate as key factors determining choices of residential location.

3. RESEARCH STRATEGY

The analytical framework in this article consists of two steps in terms of model development and empirical strategy. The first step of model development is to identify the supporting vectors (SVs) to price variances using the stepwise multi-regression approach; and the second step of empirical strategy is to forecast the housing price variances in Taipei City by employing the SVs identified by

the first step as well as other variables postulated by the hedonic price theory. Regarding the process of model development, it is first to build a typical hedonic model based on previous related theories theory (Lancaster 1966; Rosen 1974); and then to identify supporting vectors (SVs) to price directions by applying the stepwise multi-regression approach. In following this way, a SVM hedonic can be established for forecasting housing prices in Taipei city empirically.

3.1. The hedonic appraisal model

The SVM hedonic developed in this paper employs variables capturing information sets of both structural and spatial locational features within a neighborhood or sub-market. The housing structural variables include property types, floor area, land area, age of property, building height, floor level, number of bedrooms, number of living halls, number of bathrooms, and availability of a parking lot. It is important to note that sales price is usually associated with greater marketing time (Glower *et al.* 1998). Thereafter, sales duration in terms of marketing days is also incorporated into the SVM hedonic developed here. Spatial variables include transportation, accessibility, public services, aesthetic quality, and recreational facilities, the selection criteria of which are based on the literature review in Section 2.2.

After building a hedonic model based on the selection of employed information sets, it is to find out the effective hedonic variables that are useful and significant to develop the SVM model. To this end, a stepwise multiple linear regression is applied to build the hedonic model to examine and determine the significant set of variables for SVM modeling. Based on the statistical significances, variables are included or excluded from the model in order to develop the best model. Stepwise regression is the method chosen in part for its ability to reduce the risk of multicollinearity between explanatory variables.

3.2. Support Vector Machine Model

The identified SVs derived from the first step are then employed as classifiers for forecasting future market movements. SVM is favorable in modeling housing price due its ability to map the linear or non-linear input attribute space into a high dimensional feature space, and because it does not depend on the assumption of the probability distribution. For overlapping classes, given a training data set $\{(x_1, y_1),\ldots,(x_n, y_n)\}$, and $x_i \in R^n$, $y_i \in \{+1,-1\}$, Cortes and Vapnik (1995) propose a generalized optimal margin algorithm for the separating hyperplane:

$$y_i\left[\langle \boldsymbol{w}, \boldsymbol{x}_i\rangle + b\right] \geq 1 - \xi_i \quad (1)$$

Subjected to $\xi_i \geq 0, \quad i = 1,\ldots,n$.

In the case where data is inseparable in the original feature space, SVM requires mapping the input vectors into a higher dimensional feature space, and the goal of SVM is to determine the following decision function:

$$f(x) = sgn\left[\sum_{i=1}^{n} y_i \alpha_i \cdot K(x, x_i) + b\right]. \quad (2)$$

The coefficients α_i are found by solving the dual problem which maximizes the following function:

$$W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j). \quad (3)$$

With constraints, $0 \leq \alpha_i \leq C$ and $i = 1,\ldots., n$; where C is positive constant for ξ_i. In order to simplify the heavy computation of transformation, K is a kernel function that enables operations to be performed in the input space rather than the potentially high dimensional feature space. Among acceptable kernel functions for mappings are linear, polynomials, radial basis functions, and certain sigmoid functions. The study adopted an RBF kernel since it encompasses the features of a linear kernel and sigmoid kernel that are able to map linear and non-linear classifications, while maintaining lesser hyperparameters than a polynomial kernel (Hsu *et al.* 2003). The RBF kernel is given as:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp\left(-\gamma \left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|^2\right). \quad (4)$$

To avoid attributes in greater numeric ranges dominating those in smaller numeric ranges, scaling is performed on the attribute variables to transform the numeric values into the range of [0,1]. In the five-fold cross validation, samples are divided into 5 subsets, and sequentially one subset is tested using the classifier trained on the remaining 4 subsets. In the training process, SVM classification is required to identify the best numeric setting for parameters C and γ. The values of the parameter setting are determined through trial and error, and the validation of results is subjected to the accuracy requirement.

4. THE CASE STUDY

To validate the proposed approach for price forecasting, transaction records of residential properties in the adjoining twelve districts of Taipei are used in the study Preliminary findings from this step warrant the prediction accuracy of SVM.

The data time interval spans from 2007 to 2010, which encompasses a relatively short but ever-volatile temporal period that includes one market fall in late 2008 and a quick recovery in 2009.

4.1. The data

Samples are randomly drawn from the market transaction of Taipei City's properties listed in the database Gigahouse Taiwan's Real Estate Portal, within the period between 2007 and 2010. Taipei City occupies 271.8 km^2. In 2011, the city had a total of 1,089,541 households covering a total area of 178.55 km^2, with a population of 2.65 million (Taipei City Government 2012). Property types mainly comprised of two categories which are either classified as walk-up buildings or highrise apartments. The sale records in the database also furnished information related to geographical location, property structural characteristics, property amenities, and marketing information. To avoid outliers in the pricing data, only samples within ±2 standard deviations from the mean value of sale prices are included. In total 3991 samples are included in the study, and each district is represented by at least 200 samples in the analysis.

4.2. The hedonic variables

The analysis sourced data of structural attributes and locational attributes as input to predict the pricing (Table 1). A range of key variables pertaining to structural attributes in the study was extracted from the database, including property types, floor area, land area, age of property, building height, floor level, number of bedrooms, number of living, number of bathrooms, and availability of a parking lot. Amidst the effects of housing characteristics, sales price is usually associated with greater marketing time (Glower *et al.* 1998). Thereafter, the number of days housing has been on the market prior to sale has been extracted from the database and included in the analysis.

Of the three major locational attributes–namely district characteristics, transportation infrastructures, and neighbourhood amenities–distance and accessibility greatly influenced the modeling of residential location and land value. The data for locational attributes was collected by identifying the address of each residential property obtained from the property database, and further spatially linked to the distance or accessibility variables using GIS. The study adopts a Euclidean distance metric to determine the distance between two points, while the accessibility of facilities is measured by the number of facilities available adjacent to the housing within an 800-meter radius. The 800 m threshold is adopted since it is approximately equal to the 0.5 miles maximum walking distance of a commuter suggested by O'Sullivan and Morrall (1996).

The spatial dimension of Taipei City is subdivided by planning authorities into twelve administrative districts. It is well-known that real estate fixes a property in a spatial sense, and the spatial immobility defines intrinsic attributes of a dwelling, which directly affects housing quality and market value, subjected to its ease of access to public facilities and neighbourhood amenities. The variables of locational attributes shared in common by housing within the same district are labeled "district characteristics". Under this category, the key variables include district zoning, distance to central business district, and distance to the riverbank. The central business district is fixed at Taipei Main Station when measuring the distance from the property.

In terms of the effect of transportation infrastructure upon housing prices, accessibility of subway infrastructure and expressway is chosen as variables in the regression analysis. These two variables are chosen over bus stop due to the densely distributed bus stops in Taipei, rendering them insignificant as a factor influencing property value. For neighbourhood amenity, locational amenity is often articulated as the proximity of the property to public services and retail outlets (Matthews, Turnbull 2007). Our study explores potential spatial externalities by addressing a range of variables designed to capture the benefits and disadvantages of the neighbourhood and environmental characteristics of the residential property. The additional spatial variables include the number of universities, libraries, art centers, hypermarkets, department stores, supermarkets, night markets, hospitals, police stations, and fire stations located within 800m of each residential property.

The dependent variable is selling price per unit area. The advantage of using sale price per unit area in our analysis is attributed to the minimization of price range if compare to unit price. Besides, it is more common to compare property price by unit area in Taiwan. The prices have been rounded up into units of ten-thousand Taiwan dollars (TWD), and ranged from TWD 130,000 to 680,000 per unit area. For classification purposes, the selling price is disaggregated discretely into 56 classes, with most transaction prices around TWD 210–420k. Those that fall outside and above this range occur in less than 2% of the cases for each respective class.

Table 1. Variables used for structural and locational attributes

Variables	Unit
Structural attributes	
Building types *	dummy (0–walk–up building, 1–building with elevator)
Land area *	pin (about 36 square feet)
Floor area	pin (about 36 square feet)
Age *	year
number of days before sale	day
Building height	floor level
Floor level	floor level
Number of bedrooms *	number
Number of living halls *	number
Number of bathrooms *	number
Parking lot	dummy (0–no parking lot, 1–with parking lot)
Locational attributes	
District	Dummy (total 12 districts)
Distance to CBD *	meter
Distance to riverbank *	meter
Subway station *	number
Highway *	number
University *	number
Library *	number
Art center *	number
Hypermarket *	number
Department store *	number
Supermarket *	number
Night market	number
Hospital *	number
Police station *	number
Fire station *	number

* Hedonic variables selected after regression analysis.

5. RESULTS

The stepwise selection of variables offers the advantage by accommodating variables that may change in statistical significance and permits discrimination between variables, as well as their entry into and deletion from the model, thereby further reducing the risk of multicollinearity (Adair *et al.* 1996). A higher F-statistic and a threshold of significance at 0.05 are ensured for every entry of variable in the variable selection process. Furthermore, the regression models have been diagnosed to ensure no multicollinearity problems that would affect the robustness of the models. The analysis yielded 19 variables remaining after stepwise regression analysis (bulleted in Table 2). Variance inflation factor (VIF) tests have been conducted to detect multicollinearity problem of variables and it is unlikely to occur since VIF <3. The regression model has an R-squared of 64%, F-statistic of 119.12, and P-value smaller than 0.01, thus the significance of the overall model is accepted.

5.1. Accuracy of SVM

The accuracy of housing price prediction can be demonstrated by deriving a hit-rate. Based on Matysiak and Wang's (1995) findings on the correspondence between prices and valuation, the probability of valuation deviating within +/–10% is 30%, and the proportion of valuations falling within +/–20% would rise to 70%. Therefore, our SVM prediction on selling price is aimed at achieving at least a 30% and 70% hit-rate for deviation within 10% and 20% respectively.

Table 2 presents the hit-rate results of 20% allowable deviation, the optimal hit-rate 72.2%, which lies at $C = 80$, $\gamma = 2$. For similar settings, the hit-rate achieved 44.5% at an allowable deviation of 10%, with both hit-rates higher than the requirement. The selected values for parameters C and γ are adopted for the cross-validation of the remaining four subsets. After iterative SVM testing for cross-validation, the results are presented in Table 3. The hit-rate at allowable deviation of 20% resulted in 71.4%, 69.2%, 72.2%, and 70.1% respectively, which approximate the 70% requirement. The consistency of results confirms the robustness of SVM when modeling different random sample subsets.

5.2. Modeling of residential properties in Neihu and Nangang districts

The SVM model's predictability of house price in Taipei City is acceptable, but the result is less dazzling, primarily due to the locational differences of the housing area. Theoretically, the SVM model would have greater price predictability if the overall market were stratified into several homogeneous subsets. Based on such premise, a specific case study was conducted to test the SVM model predictability on the housing price for samples located only in the Neihu and Nangang districts (extracted from the 12 districts). Samples from these two neighbouring districts are chosen for their comparative proximity in location, selling prices, floor areas, and building ages. There is a total of 502 transactions recorded in these two districts. After circumventing the locational differences, the

Table 2. Hit-rate of SVM prediction within 20% deviation

	C										
γ	1	10	20	30	40	50	60	70	80	90	100
0.6	X	X	X	66.4	66.2	67.1	67.3	67.3	66.9	67.4	67.3
0.7	X	X	67.1	68.2	68.1	68.2	67.3	67.6	67.6	67.8	67.3
0.8	X	X	69.5	68.7	69.8	68.6	69.2	68.3	68.3	68.6	68.1
0.9	X	X	69.0	69.9	68.9	69.6	68.7	69.0	68.5	68.0	67.3
1	X	67.6	69.8	70.1	69.9	70.4	69.6	68.7	68.3	68.2	68.3
1.5	X	69.0	71.3	71.2	69.5	69.4	70.0	69.2	69.2	69.8	69.5
2	X	71.7	71.9	71.4	71.4	71.6	71.9	71.9	72.2	72.1	72.2
3	X	71.4	71.4	70.9	70.4	70.1	70.0	70.3	70.3	70.0	70.0
4	X	68.9	68.6	69.2	69.2	69.0	69.1	69.0	69.0	69.1	68.9
5	X	69.2	68.9	68.9	68.5	68.2	68.5	68.3	68.3	68.2	68.2
6	64.7	68.6	68.0	67.7	67.8	68.1	68.1	67.8	67.8	67.8	67.8
7	62.8	67.2	67.4	67.3	67.3	67.3	67.3	67.4	67.6	67.7	67.7
8	61.5	66.8	66.8	66.8	66.9	66.4	66.4	66.4	66.4	66.5	66.4
9	59.8	65.8	65.8	66.2	66.2	65.6	65.8	65.8	65.6	65.6	65.6
10	59.3	64.7	65.1	65.1	65.1	65.0	65.3	65.1	65.1	65.1	65.0

X: incomplete training.

SVM model hit-rate improved to 81.8% when 20% variation is allowed, while the parameters are set at $C = 50$, $\gamma = 0.5$. The improved predictability of SVM can be examined from a few perspectives.

First, there is a greater degree of locational homogeneity for the districts considered in the case study. Both Neihu and Nangang districts are sub-urban areas of Taipei City, with slower development compared to districts located closer to the city center. They have more new residential houses and higher homogeneity in district planning. In contrast, when the overall transactions in Taipei City are considered, the accuracy of SVM classification could be hampered by the imbalanced distribution of certain transportation infrastructures and neighbourhood characteristics in the city, in which the deficiency of such attributes would significantly affect the selling price in some areas. However, if the attributes are ample or overcrowded in an area, their effects on selling price would become irrelevant.

Technically, modeling housing prices in Taipei City requires a larger range of selling price classification than included in the case study. Thus, SVM modeling on Taipei City produces larger computational complexity than modeling the two districts, and therefore demand larger sample size for training set classification. However, the training samples used for the overall Taipei City SVM might be insufficient to generate a high level of predictability.

Table 3. Results of cross validation

	Cross validation				
	1	2	3	4	5
Number of testing samples	777	809	803	792	810
Number of training samples	3214	3182	3188	3199	3181
Hit-rate, %	72.2	71.4	69.2	72.2	70.1

6. CONCLUSION

The study has constructed and evaluated the usefulness of SVM classification for the mass appraisal of residential properties in Taipei City. Nevertheless, one of the key contributions in this study is the utilization and analysis of numerous hedonic variables during the preliminary stage. The hedonic model used in our study is distinct from previous studies in terms of its broad range of consideration which comprised variables of structural attributes, district characteristics, accessibility of transportation infrastructures, and neighbourhood amenities. This inclusiveness would generate a more holistic reflection of hedonic influences on the property value; furthermore, it would reduce the bias of the SVM modeling and provide better generalization.

Building upon a hedonic pricing foundation, the study devises an SVM classifier to model the effects of 19 selected variables on property price estimation. The overall accuracy of SVM estimation is comparable to Matysiak and Wang's (1995) valuation-deviation correspondence, which could achieve

an approximately 70% hit-rate for 20% allowable deviation. However, apparent improvement of SVM accuracy (82% hit-rate) has been detected when the modeling is constrained to suburban samples of Neihu and Nangang districts. In association with the difference in the degree of city development, the utility attributes of availability of public facilities and neighbourhood amenities are different across urban and suburban areas. In suburban areas that are less organized, residents might rank the importance of accessibility more highly; however, urban residents can always be compensated by other similar attributes nearby. In addition, the influences of these attributes would be more decisive in areas where their distribution is sparse rather than in areas of dense abundance. In other words, the accuracy of the SVM model would be enhanced if locational differences between the housing neighbourhoods or districts of the samples are restrained. The problem can be overcome by executing SVM modeling on distinct neighbourhoods which are distinguished from one another, or by incorporating new variables that could moderate the effects of locational differences on property value. One way to identify the locational effect is to analyze the density of locational hedonic variables within a given area. Besides, the study is designed under a situation where there is insufficient data to split training dataset from the testing dataset. Therefore, five-fold cross-validation has been adopted for the training and testing. Such approach might be biased and underestimate the true test error, however adopting five-fold cross validation is a good compensate for bias-variance tradeoff (Breiman, Spector 1992; Hastie *et al.* 2001).

In comparison with traditional mass appraisal based on multiple regression analysis, the SVM model is a superior approach which is able to circumvent problems of MRA, such as the inability to handle interacting variables, nonlinearity, and multicollinearity. The findings indicate that combining a hedonic pricing approach with SVM learning is particularly advantageous for identifying influential variables and non-linear modeling. The model is confirmed to be feasible, especially for those districts containing features that are coessential and evenly distributed in an area.

REFERENCES

Adair, A. S.; Berry, J. N.; McGreal, W. S. 1996. Hedonic modelling, housing submarkets and residential valuation, *Journal of Property Research* 13(1): 67–83. https://doi.org/10.1080/095999196368899

Adair, A. S.; McGreal, S.; Smyth, A.; Cooper, J.; Ryley, T. 2000. House prices and accessibility: the testing of relationships within the Belfast Urban Area, *Housing Studies* 15: 699–716. https://doi.org/10.1080/02673030050134565

Aye, G. C.; Balcilar, M.; Gupta, R.; Kilimani, N.; Nakumuryango, A.; Redford, S. 2014. Predicting BRICS stock returns using ARFIMA models, *Applied Financial Economics* 24(17): 1159–1166. https://doi.org/10.1080/09603107.2014.924297

Aye, G. C.; Gupta, R.; Modise, M. P. 2015. Do stock prices impact consumption and interest rate in South Africa? Evidence from a time-varying vector autoregressive model, *Journal of Emerging Market Finance* 14(2): 176–196. https://doi.org/10.1177/0972652715584267

Barari, M.; Sarkar, N.; Kundu, S.; Chowdhury, K. B. 2014. Forecasting house prices in the United States with multiple structural breaks, *International Econometric Review* 6(1): 1–23.

Beracha, E.; Wintoki, M. B. 2013. Forecasting residential real estate price changes from online search activity, *Journal of Real Estate Research* 35(3): 283–312.

Billings, S. B. 2011. Estimating the value of a new transit option, *Regional Science and Urban Economics* 41: 525–536. https://doi.org/10.1016/j.regsciurbeco.2011.03.013

Billings, S. B. 2015. Hedonic amenity valuation and housing renovations, *Real Estate Economics* 43: 652–682. https://doi.org/10.1111/1540–6229.12093

Bracke, P. 2013. How long do housing cycles last? A duration analysis for 19 OECD countries, *Journal of Housing Economics* 22(3): 213–230. https://doi.org/10.1016/j.jhe.2013.06.001

Breiman, L.; Spector, P. 1992. Submodel selection and evaluation in regression. The X-random case, *International Statistical Review / Revue Internationale de Statistique* 60(3): 291–319. https://doi.org/10.2307/1403680

Brown, J. P.; Song, H.; McGillivray, A. 1997. Forecasting UK house prices: a time varying coefficient approach, *Economic Modelling* 14(4): 529–548. https://doi.org/10.1016/S0264–9993(97)00006–0

Case, K. E.; Shiller, R. J. 1989. The efficiency of the market for single family homes, *American Economic Review* 79: 125–137.

Clark, D. E.; Kahn, J. R. 1988. The social benefits of urban cultural amenities, *Journal of Regional Science* 28: 363–377. https://doi.org/10.1111/j.1467–9787.1988.tb01088.x

Cortes, C.; Vapnik, V. 1995. Support-vector networks, *Machine Learning* 20: 273–297. https://doi.org/10.1007/BF00994018

Cristianini, N.; Shawe-Taylor, J. 2000. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801389

d'Amato, M. 2007. Comparing rough set theory with multiple regression analysis as automated valuation methodologies, *International Real Estate Review* 10: 42–65.

De Greef, I.; de Haas, R. 2000. Housing prices, bank lending, and monetary policy, *DNB Research Series Supervision Paper No. 31*.

Debrezion, G.; Pels, E.; Rietveld, P. 2007. The impact of railway stations on residential and commercial property value: a meta-analysis, *Journal of Real Estate Finance and Economics* 35: 161–180. https://doi.org/10.1007/s11146–007–9032-z

Des Rosiers, F.; Lagana, A.; Thériault, M.; Beaudoin, M. 1996. Shopping centres and house values: an empirical investigation, *Journal of Property Valuation and Investment* 14: 41–62. https://doi.org/10.1108/14635789610153461

Drake, L. 1993. Modelling UK house prices using cointegration: an application of the Johansen technique, *Applied Economics* 25(9): 1225–1228. https://doi.org/10.1080/00036849300000183

Gibbons, S.; Machin, S. 2008. Valuing school quality, better transport, and lower crime: evidence from house prices, *Oxford Review of Economic Policy* 24: 99–119. https://doi.org/10.1093/oxrep/grn008

Glower, M.; Haurin, D. R.; Hendershott, P. H. 1998. Selling time and selling price: the influence of seller motivation, *Real Estate Economics* 26: 719–740. https://doi.org/10.1111/1540–6229.00763

González, M. A. S.; Formoso, C. T. 2006. Mass appraisal with genetic fuzzy rule-based systems, *Property Management* 24: 20–30. https://doi.org/10.1108/02637470610643092

Hadjimatheou, G. 1976. *Housing and mortgage markets*. Farnborough: Saxon House.

Hamilton, S. E.; Morgan, A. 2010. Integrating lidar, GIS and hedonic price modeling to measure amenity values in urban beach residential property markets, *Computers, Environment and Urban Systems* 34: 133–141. https://doi.org/10.1016/j.compenvurbsys.2009.10.007

Hanson, S. 2004. The context of urban travel: concepts and recent trends, in Hanson, S.; Giuliano. G. (Eds.). *The geography of urban transportation*. Guilford Press, 3–29.

Hastie, T.; Friedman, J.; Tibshirani, R. 2001. *Model assessment and selection, the elements of statistical learning: data mining, inference, and prediction*. Springer New York, New York, 193–224. https://doi.org/10.1007/978–0-387–21606–5_7

Haughwout, A.; Lee, D.; Tracy, J. S.; Van der Klaauw, W. 2011. Real estate investors, the leverage cycle, and the housing market crisis, *FRB of New York Staff Report* (514).

Hendry, D. F. 1984. Econometric modelling of house prices in the United Kingdom, in Hendry, D. F.; Wallis, K. (Eds.). *Econometrics and quantitative economics*. Oxford: Blackwell.

Hess, D. B.; Almeida, T. M. 2007. Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York, *Urban Studies* 44: 1041–1068. https://doi.org/10.1080/00420980701256005

Holly, S.; Jones, N. 1997. House prices since the 1940s: cointegration, demography and asymmetries, *Economic Modelling* 14: 549–565. https://doi.org/10.1016/S0264–9993(97)00009–6

Hsu, C. W.; Chang, C. C.; Lin, C. J. 2003. *A practical guide to support vector classification*. National Taiwan University.

Kauko, T. 2003. Residential property value and locational externalities, *Journal of Property Investment & Finance* 21: 250–270. https://doi.org/10.1108/14635780310481676

Kazem, A.; Sharifi, E.; Hussain, F. K.; Saberi, M.; Hussain, O. K. 2013. Support vector regression with chaos-based firefly algorithm for stock market price forecasting, *Applied Soft Computing* 13(2): 947–958. https://doi.org/10.1016/j.asoc.2012.09.024

Keskin, B. 2008. Hedonic analysis of price in the Istanbul housing market, *International Journal of Strategic Property Management* 12: 125–138. https://doi.org/10.3846/1648–715X.2008.12.125–138

Kishor, N. K.; Kumari, S.; Song, S. 2015. Time variation in the relative importance of permanent and transitory components in the US housing market, *Finance Research Letters* 12: 92–99. https://doi.org/10.1016/j.frl.2014.11.004

Kong, F.; Yin, H.; Nakagoshi, N. 2007. Using GIS and landscape metrics in the hedonic price modeling of the amenity value of urban green space: a case study in Jinan City, China, *Landscape and Urban Planning* 79: 240–252. https://doi.org/10.1016/j.landurbplan.2006.02.013

Kryvobokov, M.; Wilhelmsson, M. 2007. Analysing location attributes with a hedonic model for apartment prices in Donetsk, Ukraine, *International Journal of Strategic Property Management* 11: 157–178.

Lamont, O.; Stein, J. C. 1999. Leverage and house-price dynamics in U.S. cities, *RAND Journal of Economics* 30(3): 498–514. https://doi.org/10.2307/2556060

Lancaster, K. 1966. A new approach to consumer theory, *Journal of Political Economy* 84: 132–157. https://doi.org/10.1086/259131

Larson, W. D. 2010. *Evaluating alternative methods of forecasting house prices: a post-crisis reassessment* [online]. Available at: http://ssrn.com/abstract=1709647 [accessed 15 November 2010]

Matthews, J. W.; Turnbull, G. K. 2007. Neighborhood street layout and property value: the interaction of accessibility and land use mix, *Journal of Real Estate Finance and Economics* 35: 111–141. https://doi.org/10.1007/s11146–007–9035–9

Matysiak, G.; Wang, P. 1995. Commercial property market prices and valuations: analysing the correspondence, *Journal of Property Research* 12(3): 181–202. https://doi.org/10.1080/09599919508724144

Mayer, C. J. 1993. Taxes, income distribution, and the real estate cycle: why all houses do not appreciate at the same rate, *New England Economic Review* May/June: 39–50.

Mayes, D. G. 1979. *The property boom*. Oxford: Martin Robertson.

Muellbauer, J.; Murphy, A. 1997. Booms and busts in the U.K. housing market, *Economic Journal* 107: 1701–1727. https://doi.org/10.1111/j.1468–0297.1997.tb00076.x

Nghiep, N.; Al, C. 2001. Predicting housing value: a comparison of multiple regression analysis and artificial

neural networks, *Journal of Real Estate Research* 22: 313–336.

Ohtake, F.; Shintani, M. 1996. The effect of demographics on the Japanese housing market, *Regional Science and Urban Economics* 26(2): 189–201. https://doi.org/10.1016/0166–0462(95)02113–2

Ortalo-Magne, F.; Rady, S. 2006. Housing market dynamics: on the contribution of income shocks and credit constraints, *Review of Economic Studies* 73: 459–485. https://doi.org/10.1111/j.1467–937X.2006.383_1.x

Osland, L. 2010. An application of spatial econometrics in relation to hedonic house price modelling, *Journal of Real Estate Research* 32(3): 289–320.

O'Sullivan, S.; Morrall, J. 1996. Walking distances to and from light-rail transit stations, *Transportation Research Record: Journal of the Transportation Research Board* 1538: 19–26. https://doi.org/10.3141/1538–03

Rosen, S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition, *Journal of Political Economy* 82(1): 34–55. https://doi.org/10.1086/260169

Shin, K.; Washington, S.; Choi, K. 2007. Effects of transportation accessibility on residential property values: application of spatial hedonic price model in Seoul, South Korea metropolitan area, *Transportation Research Record* 1994(1): 66–73.

Sirpal, R. 1994. Empirical modeling of the relative impacts of various sizes of shopping centers on the values of surrounding residential properties, *Journal of Real Estate Research* 9: 487–505.

Smith, L. B.; Ho, M. H. C. 1996. The relative price differential between higher and lower priced homes, *Journal of Housing Economics* 5(1): 1–17. https://doi.org/10.1006/jhec.1996.0001

Song, Y.; Sohn, J. 2007. Valuing spatial accessibility to retailing: a case study of the single family housing market in Hillsboro, Oregon, *Journal of Retailing and Consumer Services* 14: 279–288. https://doi.org/10.1016/j.jretconser.2006.07.002

Taipei City Government. 2012. *Taipei City statistical abstract 2011*. Taipei City, Taiwan.

Waddell, P.; Berry, B. J.; Hoch, I. 1993. Residential property values in a multinodal urban area: New evidence on the implicit price of location, *Journal of Real Estate Finance and Economics* 7(2): 117–141. https://doi.org/10.1007/BF01258322

Wang, X.; Wen, J.; Zhang, Y.; Wang, Y. 2014. Real estate price forecasting based on SVM optimized by PSO, *Optik-International Journal for Light and Electron Optics* 125(3): 1439–1443. https://doi.org/10.1016/j.ijleo.2013.09.017

Wheaton, W. C.; Chervachidze, S.; Nechayev, G. 2014. Error correction models of MSA housing 'supply' elasticities: implications for price recovery, *MIT Department of Economics Working Paper No. 14–05*. https://doi.org/10.2139/ssrn.2382920

Whitehead, C. M. E. 1974. *The UK housing market: an econometric model*. Saxon House, Farnborough.

Wu, J.; Deng, Y.; Liu, H. 2014. House price index construction in the nascent housing market: the case of China, *Journal of Real Estate Finance and Economics* 48(3): 522–545. https://doi.org/10.1007/s11146–013–9416–1

Wu, L.; Brynjolfsson, E. 2013. *The future of prediction: how Google searches foreshadow housing prices and sales*. Available at SSRN: https://doi.org/10.2139/ssrn.2022293

Zhang, H.; Yang, F.; Li, Y.; Li, H. 2015. Predicting profitability of listed construction companies based on principal component analysis and support vector machine–evidence from China, *Automation in Construction* 53: 22–28. https://doi.org/10.1016/j.autcon.2015.03.001

Zurada, J.; Levitan, A.; Guan, J. 2011. A comparison of regression and artificial intelligence methods in a mass appraisal context, *Journal of Real Estate Research* 33: 349–387.