INCORPORATING STEIN ESTIMATOR AND SEMIPARAMETRIC METHODS FOR ENHANCING HOUSING INDEX

CHENG CHIN TIONG
Universiti Teknologi Malaysia

JANICE LEE YIM MEI
Universiti Teknologi Malaysia
and
HISHAMUDDIN BIN MOHD ALI
Universiti Teknologi Malaysia

ABSTRACT

The rapid growth of Malaysia's property leads to transformation of physical property investment into financial products such as housing derivative or home equity insurance. It stimulates the demand of more reliable, accurate, stable and updated index. The instability in calculating housing index is unfavourable to the development of markets for housing derivatives. Previous researcher claimed that most of the researchers using geospatial data to model housing market rather than development of housing index and the official housing index that included geospatial data in hedonic price index are limited such as FNC Residential Price Index and RPData-Rismark Daily Home Value Index. Therefore, it is important to explore the role of spatial model in developing housing index since it had increased the reliability in predicting house price according to few researchers. Two regressions modelling techniques such as stein estimator and semiparametric methods will be investigated in order to yield a consistent housing index since it can enhance the predictive power in property value by using hedonic model. The study area will focus on double storey terrace houses located in Plentong, Johor Bahru, Malaysia. The first objective is to explore the effectiveness of geocoded data (longitude, latitude) in hedonic model to produce housing index. The second objective is to identify the enhancement of hedonic model by using stein estimator and semiparametric methods in developing housing index. This study found that stein estimator could enhance the development of housing index in the level of index revision and mean squared error.

Keywords: housing derivative, housing price index, geospatial data, stein estimator, semiparametric, hedonic model

INTRODUCTION

The housing market has formed one of the largest assets in Malaysia. The boom and bust in the housing market will direct impact toward economy in Malaysia. Thus, transacted price that is based on supply and demand becomes an indicator to determine the current environment of the housing market. Investors and government in Malaysia can utilize it to construct their investment strategies or housing policy. Moreover, the rapid growth of property leads to transformation of physical property investment into financial products such as housing derivative or home equity insurance. It stimulates the demand of more reliable, accurate, stable and updated index.

The purpose to conduct this research is to enhance the reliability and stability of the housing index in Malaysia. Therefore, it can transform into tradable index for financial products in the future. The reliable index plays a main role in facilitating the transaction of housing derivatives. Clapham (2006) found that hedonic model would be suitable to assist the development of housing derivatives index compare to repeated sales and median index because it is less prone to index revision. Although Malaysia Housing Price Index (MHPI) is constructed from hedonic model, but it may be improved by considering spatial elements and regression modelling techniques in development of hedonic model.

The model misspecification is one of the problems in hedonic model to construct housing index. The hedonic model may suffer from omitted variables, especially the effect of location for housing value. In specifically the neglected from spatial and temporal dependences effects will induce bias in coefficient estimates and/or heteroskedasticity issues (Nappi-Choulet and Maury 2009). Apart from that, there are previous studies stated the significant effects of spatial elements toward the predictive power of hedonic models in pricing a property (Pace *et al.*1998; Can and Megbolugbe 1997). In development of housing index, Song and Wilhelmsson (2010) claimed that the inclusion of spatial elements leads to improvement of hedonic in term of predictive power but it does not affect housing index significantly for condominium. But, different types of property may yield different results therefore it is important to determine the effects of spatial dependency and heterogeneity toward development of housing index for double-storey house in Plentong Johor.

Apart from that, regression modelling also employed estimator to estimate the marginal prices for each variable in hedonic pricing model. The common estimator that widely used in hedonic model for housing price index is known as ordinary least square (OLS). Bao and Wan (2007) pointed out that although OLS estimator is the most popular technique in hedonic pricing analysis, but it suffered from the problem of model imprecision if relevant information is not contained in the data. This is because the conventional least square does not allow the utilization of non-sample information such as expert opinions. Therefore, it may induce multicollinearity of data and result in poor and large variance estimation. Thus, the introduction of stein rule estimator is useful to handle collinearity problem of data in previous studies. Knight *et. al.* (1993) claimed that stein rule estimation is useful for the valuation problem encountered by appraisers and assessors particularly handling collinearity and non sample information. According to Bao and Wan (2007) traditional stein rule estimation is essentially a weighted average of the OLS and restricted least squares estimators, with weights that are functions of the computed F statistic used to test the set of hypotheses regarding the restrictions representing the extraneous information. Bao and Wan (2007) conducted a study by using a stein rule estimation technique to handle collinearity problem. They found that this estimator can improve the hedonic price valuation. Moreover, the domination of stein rule estimation over ordinary least square (OLS) estimator in term of mean square error are stated in previous studies (Namba and Ohtani 2012;Knight and Hill 1992; Knight and Hill 1993). In past studies, most of the hedonic models to construct index are estimated by using OLS (Goh *et. al.* 2012; Costello and Watkins 2002; Clapham *et. al.* 2006). In order to enhance the predictive power of hedonic model adaptation of stein rule estimation in developing an accurate housing index is possible.

Moreover, functional form misspecification may affect on the accuracy of hedonic model in predicting housing index. According to Bao and Wan (2004) if wrong functional form is selected to compute hedonic model the validity of interpreting the estimated parameters will be highly questionable. Generally, semi-log as a famous functional form proposed by Box and Cox (1964) and widely used in constructing housing index (Goodman 1978; Halvorsen and Pollakowski 1981; Maurer *et. al.* 2004; Clapham *et. al.* 2006;Widlak and Tomczyk2010). Unfortunately, Coulson *et. al.* (2008) mentioned that there is risk of functional form misspecification by using the semi - log model. Thus, in order to avoid this problem some researchers are using non parametric or semiparametric instead of semi-log model. In order to avoid functional form misspecification the adoption of nonparametric or semiparametric regression could be an alternative choice to develop hedonic model. Breiman (2001) point out that semiparametric and non parametric approaches are representative of an algorithmic modelling culture. Both methods are suitable for many hedonic modelling situations where incomplete knowledge prevents the exact a priori specification of non linear or non stationary components of functional form (Hannonen 2006). Some studies had shown that nonparametric or semiparametric method used for hedonic price model fit the data better than parametric models in term of more accurate out of sample predictions (Pace 1993; Messe and Wallace 1991; Pavlov 2000; Clapp 2004). However, as mentioned by Hill (2013) there is a lack of discussion on nonparametric method in constructing housing index. Clapp (2004) conduct on semiparametric hedonic model in developing housing index and the model can reduce out of sample mean squared error of 11% when compared to the OLS. Due to lack of study in this area the application in housing index using semiparametric or non parametric is still questionable. Therefore, adopt semiparametric methods in developing housing index is another possible study suggested by Bao and Wan (2004).

The first objective in this study is to explore the effectiveness of geocoded data (latitude and longitude) in hedonic model to produce housing index. The second objective is to identify the enhancement of hedonic model to construct housing index by using stein estimator and semiparametric methods.

HEDONIC REGRESSION METHOD

The hedonic method recognizes that a property is a composite product that can be described by their attributes or characteristics. The attributes are not sold separately, yet regressing the attributes on the sales price of the composite product yields the marginal contribution of each attribute to the sales price (Rosen 1974). Haan and Diewert (2011) claimed that the demand and supply for the properties implicitly determine the characteristics' marginal contributions to the prices of the properties. Regression techniques can be used to estimate those marginal contributions or shadow prices. The main advantage of hedonic regression method is to provide quality adjusted price indices.

Haan and Diewert (2011) mentioned that the property price P_n^t of property n in period t is a function of a fixed number of K characteristics measured by "quantities" (Z_{nk}^t). The T+1 time periods going from the base period 0 to period T. Equation 1 and 2 indicated standard regression techniques which are parametric model used to estimate the marginal contributions of each characteristic

$$P_n^t = \beta_0^t + \sum_{k=1}^{k} \beta_k^t Z_{nk}^t + \epsilon_n^t \quad (1)$$

$$\ln P_n^t = \beta_0^t + \sum_{k=1}^{k} \beta_k^t Z_{nk}^t + \epsilon_n^t \quad (2)$$

The characteristic parameters β_k^t in 1 and 2 are allowed to change over time. It is an idea that housing market conditions determine the marginal contributions of the characteristics. Due to demand and supply condition changes there is no a priori reason to expect that these contributions are constant (Pakes 2003).

HEDONIC TIME DUMMY METHOD IN CONSTRUCTING HOUSING INDEX

The time dummy variable approach used to construct a quality adjusted price index in academic studies. The equation 3 for time dummy variable approach is stated as below. Running one overall regression on the pooled data of the samples S(0), S(1),…S(T) relating to periods t=0,…,T (with sizes N(0), N(1),…,N(T)) yields coefficients β^0, δ^t (t=1,…T) and β_k (k=1,….,K). The time dummy parameter shifts the hedonic surface upwards or downwards and measures the effects of "time" on the logarithm of price. Exponentiating the time dummy coefficients thus

^

controls for changes in the quantities of the characteristics and provides a measure of quality-adjusted house price change between the base period 0 and each comparison period t. Therefore, the time dummy index going from period 0 to period t is shown in equation 4. This simple hedonic model can be further developed into a spatial hedonic model that incorporated with spatial elements (longitude and latitude) for developing housing index.

$$\ln P_n^t = \beta_0 + \sum_{t=1}^{T} \delta^t D_n^t + \sum_{k=1}^{k} \beta_k Z_{nk}^t + \epsilon_n^t \ (3)$$

$$P_{TD}^{0t} = \exp\left(\delta^t\right) \ (4)$$

STEIN RULE ESTIMATION

Stein rule estimator is based on the linear regression model that can model the house price. It consists of both dependent (sale price of the house) and independent variable (explanatory variable such as the housing characteristic). The coefficient in a model is representing the marginal contributions b_i of housing characteristics to the sale price of price. Random error $\boldsymbol{\epsilon}$ is a random error representing factor not appearing explicitly in the model. This technique provides the best unbiased estimates if the errors are normally distributed. In term of property valuation appraisers often have information about the regression coefficients. For example, appraiser may know that changes in particular characteristics add to property value or that other changes may reduce values. Alternatively, an appraiser may be sceptical that a certain characteristic has influence on price or may have prior guess as to the marginal price of a characteristic. Each of these prior, subjective or personal assumptions can be called non-sample information.

$$\widehat{\beta_i} = \lambda b_i^* + (1-\lambda) b_i \ (5)$$

Stein estimator is a shrinkage estimator. According to Schafer and Strimmer (2005) linear shrinkage approach used to combine both estimators in a weighted average as shown in equation 5. The λ represent a shrinkage intensity. It can ranged from 0 to 1. By selecting the optimal shrinkage intensity it can lead to better prediction by reducing the mean squared error.

SEMIPARAMETRIC REGRESSION

Obviously, linear parametric models tend to be somewhat restrictive. Therefore, the need of arises for greater flexibility in the specification of a regression equation such as non linear relationships for regression function. By combining non parametric and parametric regression it yield semiparametric regression model or partial linear model (Panik 2009). The simple model below is modified from Clapp (2004). The model below consisted of linear characteristics of housing characteristic (X_i), time dummy (D_n^t) toward house price and non linear characteristics of coordinates (longitude and latitude).

$$\ln\text{Price} = X_i\beta_1 + \delta_t D + f(\text{lon,lat}) + \varepsilon \ (6)$$

THE VARIABLES IN HEDONIC MODELS

The selection of variables is important to develop a reliable hedonic model. In this study house price of the double storey terraced house acts as dependent variable. After that, it followed by independent variables such as time, land area, built-up area, types of property, tenure and numbers of bedroom. According to Hamid (2007) land area of house contributes positively by 35% toward house price. The reason is bigger land size will provide a variety of usage for homeowners such as a larger car porch. In addition, larger built-up area also provides additional space to accommodate more people in a house. Thus, it has a positive effect on house price (Rodriguez and Sirmans 1994; Carroll *et. al.* 1996). The numbers of bedrooms in a house has a positive relationship toward house price according to previous researches (Fletcher *et. al.* 2000; Choy *et. al.* 2007). In Malaysia, the ownership of land is categorized into leasehold and freehold. Leasehold properties will be reverting back to state government in a period of time while the freehold property will enjoy a perpetual interest. Therefore, the advantage of freehold property will lead to higher price theoretically. Furthermore, Malaysia consists of two types of lot which are Malay reserve lot and International lot. Generally, the Malay reserve lots will be offered at a cheaper price compared to International lot. The transaction for Malay reserve lots is rather restrictive compare to international lots according to Malay Reservation Land Law in Malaysia. Therefore, less restriction for international lots in transaction lead to higher price compare to Malay reserve lots. Types of property such as low and medium cost properties also contribute to house price. According to Seventh Malaysia Plan (1996-2000) and Eight Malaysia Plan (2001-2005) will provide affordable houses for low income group in Malaysia. Furthermore, private sectors are compulsory to construct 30% low cost housing in every residential development which monitored by local authority. These two types of properties are cheaper than ordinary properties. It may probably due to different design and materials.

METHODOLOGY

This section comprised of a traditional hedonic model (OLS), spatial hedonic model (SHM) and semiparametric hedonic model (SPM) that used to construct a housing index in this study.

TRADITIONAL HEDONIC MODEL (OLS)

The traditional hedonic model is a simple hedonic model that only included housing characteristics in constructing a quality adjusted housing index. The location of each house is not included in this model.

$$\ln Price_{it} = \beta_0 + \delta D_t + \beta_1 Land_i + \beta_2 Built_i + \beta_3 Tenure_i + \beta_4 Lot_i + \beta_5 Housetype_i + \beta_6 Bedrooms_i + \beta_7 Category_i + \varepsilon_{it} \quad (7)$$

$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$ = coefficients for respective parameter

δ = coefficient for monthly time dummy variable (one if t in month D, otherwise zero)

Land = Land Area of a house in square metres

Built = Built up area of a house in square metres

Tenure = Tenure of a house in dummy variable (one if freehold, otherwise zero)

Lot = Position of a house in dummy variable (one if corner or end lot, otherwise zero)

Bedrooms = Number of Bedrooms of a house

Category = Category of land for a house in dummy variable (one if Malay Reservation Land (MRL), otherwise zero)

Housetype = Categorize house into low cost(1), medium cost(2), and ordinary house type(3).

$\ln Price_{it}$ = logarithm form of sale price of house i at time t

i, t = Number of transactions in house i at time t

The traditional hedonic model above is produced by using ordinary least square. The sample of data from the based period (January in the year 2006 for this study) will combine with a sample of data from respective month (for an example February in the year 2006) to perform regression analysis. Housing index for February in the year 2006 is obtained through the exponential of coefficient for monthly time dummy variable. This method is repeated to produce subsequent periods of housing index. In this study it contained 71 models to produce 71 periods of housing index.

SPATIAL HEDONIC MODEL (SHM)

The Spatial Hedonic Model uses trend surface analysis to extend the traditional hedonic model. It has been transformed by assigning polynomial expansion of absolute location variables (x and y coordinates) into the traditional hedonic model. According to Bitter *et. al.* (2007), this is possible due to the employment of polynomial expansion by raising variables to successive powers. This method allows parameter estimates to vary over space (Jones and Cassette 1992) and produces location dependent pattern of the sale price over two or three dimension surfaces (Xu 2008). Clapp (2004) stated the equation in developing housing index as equation 8.

$$\ln Price = \beta_0 + \delta D_t + \beta_1 Land_i + \beta_2 Built_i + \beta_3 Tenure_i + \beta_4 Lot_i + \beta_5 Housetype_i + \beta_6 Bedrooms_i + \beta_7 Category_i + \gamma_1 lat_i + \gamma_2 lon_i + \gamma_3 Lat_i^2 + \gamma_4 Lon_i^2 + \gamma_5 LatLon_i + \varepsilon_{it} \qquad (8)$$

$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5$ = coefficients for respective parameter

δ = coefficient for monthly time dummy variable (one if t in month D, otherwise zero)

Land = Land Area of a house in square metres

Built = Built up area of a house in square metres

Tenure = Tenure of a house in dummy variable (one if freehold, otherwise zero)

Lot = Position of a house in dummy variable (one if corner or end lot, otherwise zero)

Bedrooms = Number of Bedrooms for a house

Housetype = Categorize house into low cost(1), medium cost(2), and ordinary house type(3)

Category = Category of land for a house in dummy variable (one if Malay Reservation Land (MRL), otherwise zero)

lat = Latitude of a house

lon = Longitude of a house

$\ln Price_{it}$ = logarithm form of sale price for house i at time t

i, t = Number of transactions for house i at time t

The spatial hedonic model above produced by using ordinary least square is named as SHM. The spatial hedonic model yield by stein/shrinkage estimator (from the "corpcor" function in R package) is named as stein in this study . The way to compute individual housing index is similar to traditional hedonic model above. Once vector or coefficient of time dummy variable is obtained from regression it can be translated into price index by simply exponentiating it.

SEMIPARAMETRIC IN SPATIAL HEDONIC MODEL (SPM)

This section illustrated the computation of spatial hedonic model by using semiparametric method. As mentioned above semiparametric is a combination of parametric and non parametric regression to yield a model. In this study spatial element that exerts non linear characteristics is categorized as non parametric components. The spatial hedonic model computed by semiparametric method is written as below:

$$\ln Price = \beta_0 + \delta D_t + \beta_1\, Land_i + \beta_2\, Built_i + \beta_3\, Tenure_i + \beta_4\, Lot_i + \beta_5\, Housetype_i + \beta_6\, Bedrooms_i + \beta_7\, Category_i + f(lat,lon) + \varepsilon_{it} \quad (9)$$

$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$ = coefficients for respective parameter

δ = coefficient for monthly time dummy variable (one if t in month D, otherwise zero)

Land = Land Area of a house in square metres

Built = Built up area of a house in square metres

Tenure = Tenure of a house in dummy variable (one if freehold, otherwise zero)

Lot = Position of a house in dummy variable (one if corner or end lot, otherwise zero)

Bedrooms = Number of Bedrooms of a house

Category = Category of land for a house in dummy variable (one if Malay Reservation Land (MRL), otherwise zero)

Housetype = Categorize house into low cost(1), medium cost(2), and ordinary house type(3)

$\ln Price_{it}$ = logarithm form of sale price for house i at time t

i, t = Number of transactions for house i at time t

f(lat,lon) = Non-parametric components (latitude and longitude)

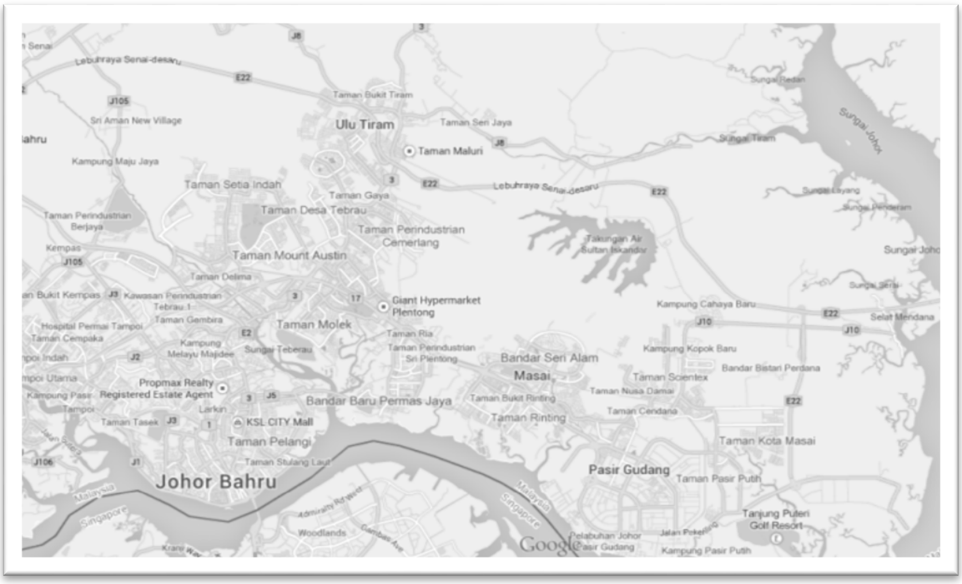
A semiparametric computation of spatial hedonic model above is performed by using penalized spline smoothing (from “spm” function in R package). The way to produce a housing index is similar to the hedonic models in previous section whereby exponentiating the coefficient of time dummy variable.

SAMPLE OF DATA

The sample is collected from case study Plentong which is a town under the district of Johor Bahru. Plentong is a town that comprised of several residential areas. There are Bandar Bistari

Perdana, Bandar Baru Kota Putri, Bandar Baru Permas Jaya, Bandar Baru Seri Alam,Desa Harmoni, Taman Air Biru, Taman Bukit Dahlia, Taman Cahaya Masai, Taman Desa Cemerlang, Taman Desa Harmoni, Taman Ehsan Jaya, Taman Desa Jaya, Taman Flora Height, Taman Iskandar, Taman Johor Jaya, Taman Kota Masai, Taman Maju Jaya, Taman Mawar, Taman MegahRia, Taman Molek, Taman Nora, Taman Nusa Damai, Taman Pasir Emas, Taman Pasir Puteh, Taman Pelangi, Taman Plentong Baru, Taman Ponderosa, Taman Ria, Taman Rinting, Taman Scientex, Taman Sentosa, Taman Sri Tebrau and Taman Sri Tiram. Plentong is located at the west part of Johor Bahru city centre and surrounded by flagship A and D of Iskandar region which had shown in Figure 1. Flaghip A is located at Johor City Centre that act as southern gateway development where involved key economic such as financial and urban tourism. Flagship D is located at eastern gate development which is Pasir Gudang and it acts as a manufacturing hub in southern region of Malaysia. The advantage of using the Plentong town as study area is because it provides sufficient housing transaction, particularly in a double storey terrace house in developing models. It is because the housing index has to be constructed in most disaggregated level, which is a monthly index in this case. The time frame is started from year 2006 until 2011. So, housing performance before and after Global Financial Crisis 2008 for specific area which is Plentong can be observed through housing index.

Figure 1: Location Map of Plentong



(Source: Google Map)

This study constructs a hedonic housing index using property transaction from the Malaysia Department of Valuation and Property Services. In order to construct a reliable housing index this study will consider on two issues which are geographical boundary and types of property. The problem of geographical aggregation might cause sample selection bias due to the different socioeconomic structure for each locality. Therefore, this study selects Plentong town where under the district of Johor Bahru, Malaysia. In addition, this housing index is particularly developed for double-storey terrace house. The selection of one type of residential property could minimize the heterogeneity between different types of property that may affect the reliability of the model. The monthly housing index is constructed from year 2006-2011.

DESCRIPTIVE ANALYSIS

Table 1: Double Storey Houses in Plentong Transacted From Year 2006-2011

	MINIMUM	MEDIAN	MEAN	MAXIMUM
Lnprice (MYR)	8.666	11.813	11.694	13.43
Land (sq.metres)	58.52	130.06	129.6	602.97
Built (sq.metres)	46.22	105.64	105.92	304.62
Tenure	0	1	0.9023	1
Lot	0	0	0.1275	1
Housetype	1	3	2.314	3
Bedrooms	0	3	2.86	5
Lat	1.439	1.504	1.511	1.595
Lon	103.5	103.8	103.8	104.2
Category	0	0	0.244	1

(Lnprice =ln(price), Land = Land area, Built=Buit up area, Tenure= Leasehold (0)& Freehold(1), Category= MRL(1) & International lot(0), Housetype = ordinary(3), medium(2) and low cost(1), Bedrooms= number of bedrooms, Lot=Corner/End (1)&Intermediate(0), Lat= Latitude, Lon=Longitude, Total sample in use= 4,621)

Table 1 indicated output of descriptive analysis from the sample used to create hedonic models. The land area and built up area are measured in square metres while tenure, lot and category are in the form of dummy variables. Housetype used number to categorize ordinary (3), medium (2) and low cost (1) double terrace house. Bedroom is recorded in number form. The latitude and longitude are coordinates for a house. The house price (MYR) is transforming into the logarithm form (Lnprice). The maximum and minimum of Lnprice is equal to 13.43 and 8.666 respectively. The average value of Lnprice in this sample is 11.694. Furthermore, the average land area for Plentong double terrace house is 129.6 square metres. The land areas of terrace houses in Plentong area reach to maximum 602.97 square metres. The mean and median built up area of a double storey terrace house in this sample is 129.6 square metres and 130.06 respectively. The mean value near to one in tenure indicated that houses in Plentong majority holding freehold titles. Besides that, intermediate lot is majority in constructing a housing index. The analysis had identified that most of double storey houses transacted are medium cost houses from the year 2006-2011 according to the mean value by 2.314. The statistic had shown that each house in Plentong has three bedrooms in average. In the Plentong area, 24.4 % land is Malay Reserve lots and remaining is international lots.

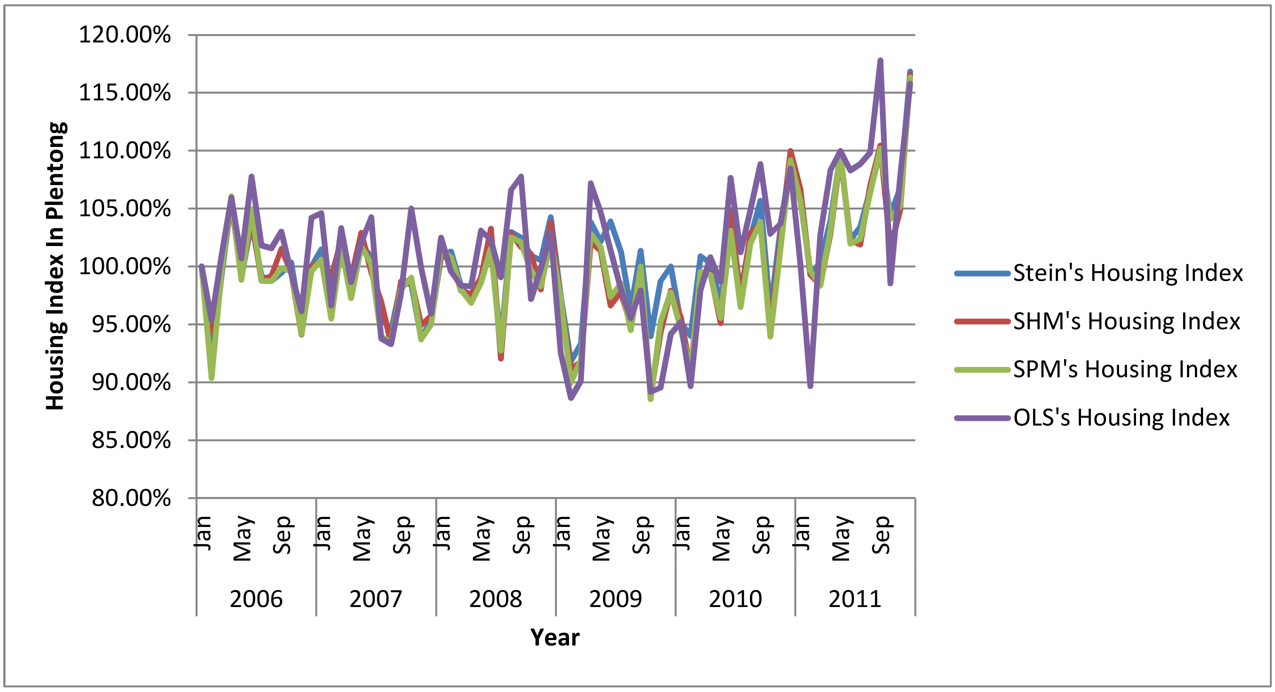
PREDICTIVE POWER OF VARIABLES

Table 2: Significance of Variables in 71 Hedonic Models

p-value	0.0001	0.001	0.01	0.05	overall
Time	0	0	6	1	9.86%
Land (sq.metres)	20	16	20	4	84.51%
Built (sq.metres)	39	16	11	2	95.77%
Tenure	0	0	4	1	7.04%
Lot	0	1	3	2	8.45%
Housetype	47	15	4	2	95.77%
Bedrooms	0	8	10	12	42.25%
Lat	0	0	9	9	25.35%
Lon	0	0	5	5	14.08%
Lat2	1	1	4	2	11.27%
Lon2	0	0	5	5	14.08%
Latlon	0	0	10	8	25.35%
Category	0	0	1	5	8.45%
Adjusted R square	mean	0.878372			
	max	0.9224			
	min	0.7462			

The selection of significant variables to construct hedonic models would be the initial step before housing index. This study consisted of 71 models in order to analyze their significance in hedonic model. Obviously, 95.77%, 95.77% and 84.51% of all models has been test shown that housetype, land area and built up area has strong predictive power toward the house price. After that, 42.55% of 71 models shown that number of bedrooms significant in predicting house price. Time effect, category, tenure and lot imposed weaker predictive power toward house price as indicated by the table above. Although they are weak but it cannot be removed from hedonic model. This is because it can still able to increase the predictive power of models. Furthermore, spatial elements (lat, lon, lat2, lon2, latlon) are important indicators to predict house price as shown by the analysis above and it should be included in a model to construct housing index. The reliability of 71 models is acceptable because the adjusted R square is 87.83% on average for each model by refer to previous research (Clapham *et. al.* 2006, Clapp 2004). There are some models recorded maximum predictive power with adjusted R square 0.9224. To conclude, these models are suitable to construct a housing index using the hedonic model.

Figure 2: Housing Index in Plentong



The housing index is constructed by Stein estimation, Spatial hedonic model, Semiparametric model and Ordinary Least Square model. Stein estimation, Spatial hedonic model and Semiparametric model construct a housing index by including time effect, housing characteristic and coordinates of each house and Ordinary Least Square model only contained time effect and housing characteristic. The line chart above depicts the performance of double storey terrace houses in Plentong town from year 2006 until 2011. The fluctuation of housing performance from year 2006 until 2011 is stable. Furthermore, from year 2006 until 2009 the housing performance is on the downtrend. This scenario may be affected by Global Financial Crisis year 2008. After that, it starts to rebound at year 2010. This trend is ascertained four housing index that constructed from different methods. The volatility of each housing index is determined by standard deviation. OLS's housing index, Stein estimation, Spatial hedonic model and Semiparametric model records 6.04, 4.51, 4.78 and 4.91 respectively. Therefore, OLS's housing index tends to be more volatile compare to another three housing index. The high volatility may drive by model misspecification in term of spatial elements such as coordinates.

OUTSAMPLE TEST

The outsample test will disaggregate 10% number of transactions from a hedonic model. After that, the remaining 90% number of transactions will construct hedonic model to predict the individual value of the 10% number of transactions. The difference between real value and predictive value will be squared and sum up. Then it divided by its' 10% number of transactions. The results for four different hedonic models are shown as below.

Table 3: Mean Square Error (MSE) Among Four Models

Hedonic Models	MSE
OLS	0.082175
SHM	0.070582
SPM	0.070016
Stein	0.069031

(OLS=Ordinary least square (without location variables), SHM=Spatial Hedonic Model, SPM=Semiparametric regression model, Stein=Stein estimation)

Table 3 indicated that housing index that developed from OLS imposed a larger MSE compare to SHM, SPM and Stein which is 0.082175. The result above had shown that Stein estimation recorded a lowest MSE or in other word it produces a hedonic model with greater predictive power compare to OLS, SHM and SPM. Although both SPM and SHM included spatial elements in their hedonic models, but the MSE from SPM is smaller than SHM. The result agreed with (Bao and Wan 2004) that the semiparametric model and the stein estimation model has better predictive power compare to OLS model.

CONSISTENCY TEST

This test will evaluate the housing index develop by four hedonic models using only 50% number of transactions. The difference of housing index constructed between full transaction and 50% numbers of transaction will be accumulate (taking the absolute value) and divided by 71 periods to obtain their level of index revision.

Table 4: Level of Index Revision Among Four Models

Types	Level of Index Revision
Stein's Housing Index	±3.82%
SPM's Housing Index	±4.49%
OLS's Housing Index	±5.29%
SHM's Housing Index	±6.27%

(OLS=Ordinary least square (without location variables), SHM=Spatial Hedonic Model, SPM=Semiparametric regression model, Stein=Stein estimation)

Obviously, Housing index constructed by Stein estimation obtains a lower level of index revision among the models above which is ±3.82%. Semiparametric method produces a housing index with ±4.49% of index revision which is second lowest among these models. The housing index developed from OLS and SHM is revised ±5.29% and ±6.27% on average for each periods. Although OLS produced a lower level of index revision compare to SHM, but it has higher MSE which is questionable in developing a housing index. Therefore, this study suggests that Stein estimation is a good estimator to develop housing index in term of MSE and level of index revision.

Figure 3: Stein's Housing Index

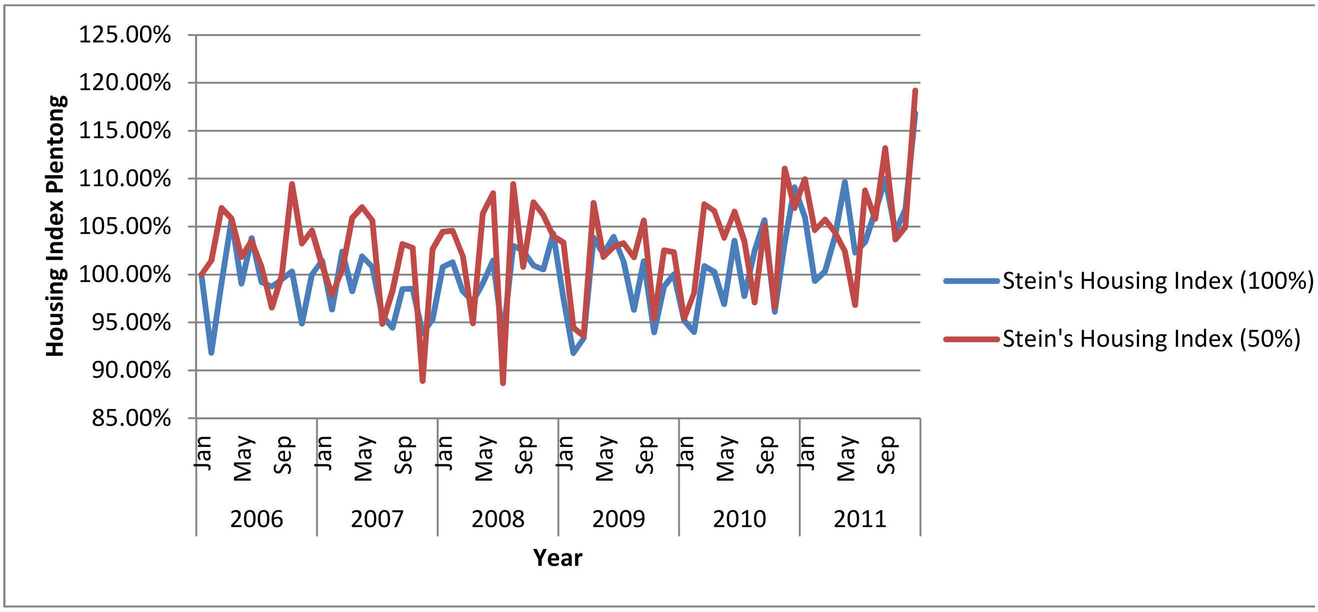


Figure 2 indicated Stein's housing index developed from two different numbers of transactions. Although Stein's housing index with 50% number of transaction is deviate from Stein's Housing Index with 100% number of transaction but both move in the same direction.

Figure 4: SPM's Housing Index

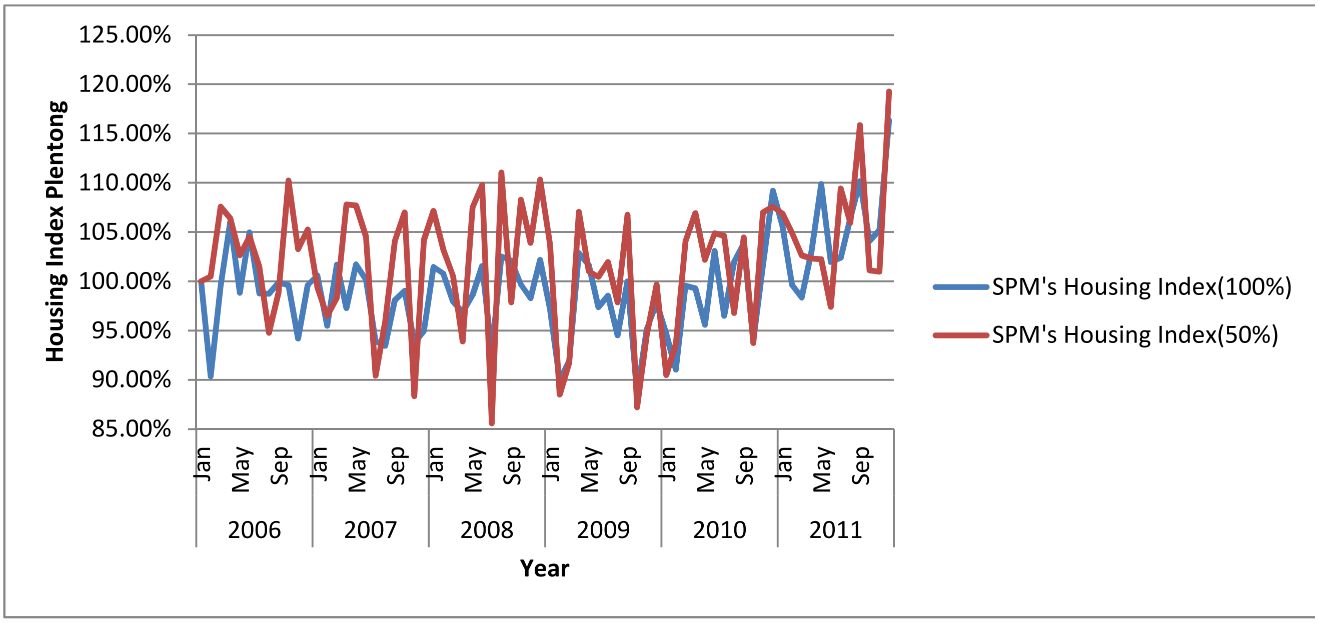


Figure 3 indicated SPM's housing index developed from two different numbers of transactions. Although SPM's housing index with 50% number of transaction is deviate from SPM's Housing Index with 100% number of transaction but both move in the same direction.

Figure 5: OLS's Housing Index

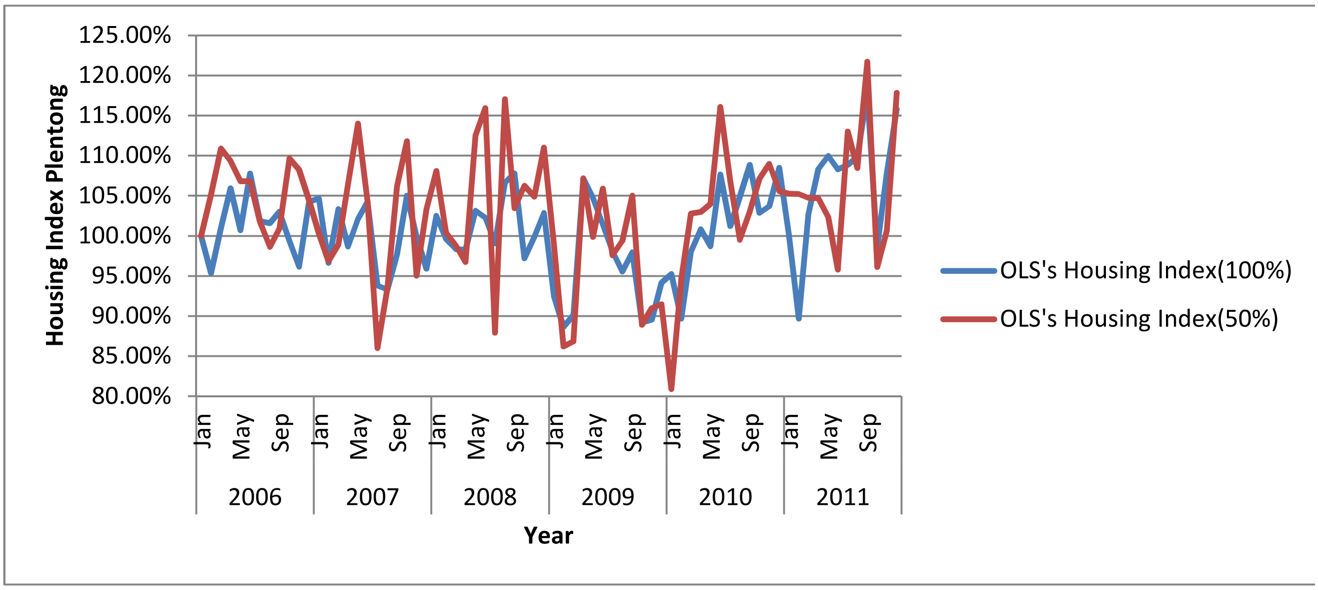


Figure 4 indicated OLS's housing index developed from two different numbers of transactions. Although OLS's housing index with 50% number of transaction has same pattern with OLS's Housing Index with 100% number of transaction but it imposed greater fluctuation compared with Stein's and SPM's housing index with 50% number of transaction.

Figure 6: SHM's Housing Index

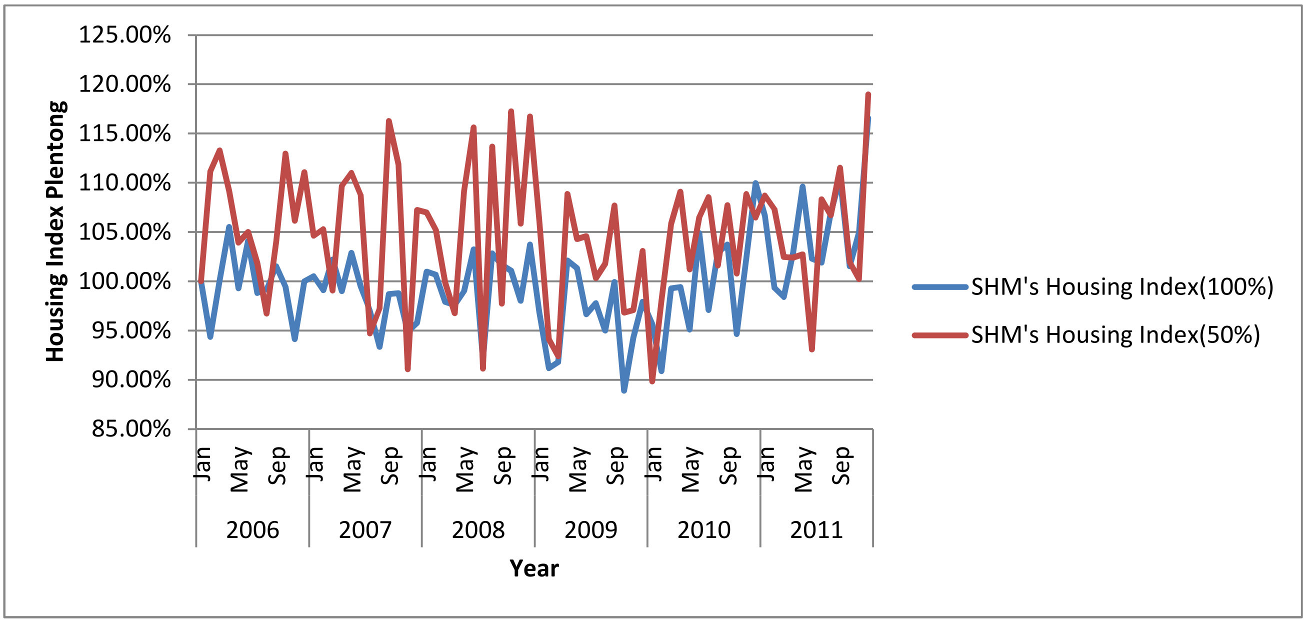


Figure 4 indicated OLS's housing index developed from two different numbers of transactions. SHM's housing index with 50% number of transaction has different pattern with SHM's Housing Index with 100% number of transaction. The deviation of SHM's housing index with 50% number of transaction from it full sample housing index is great.

CONCLUSION

Housing index is a very important indicator to determine the housing performance. It is useful in formulating housing policy or investment strategies. Furthermore, transformation of physical performance into housing derivative will require a more precise and accurate housing index. Therefore, in this study has achieved two objectives in enhancing housing index. First, location is an important variable to enhance hedonic model in terms of their mean square error. The second objective is stein estimation dominates ordinary least square estimation and semiparametric regression in constructing housing index according to level of index revision. Therefore, stein estimation is not only effective in predicting house price but also consistent in producing housing index.

REFERENCES

Abdul Hamid Mar Iman (2007). Combining Geographic Information System and Regression Model to Generate Locational Value Residual Surfaces in the Assessment of Residential Property Values. *Pacific Rim Property Research Journal*. 13(1), 35-62.

Bao, H. X., & Wan, A. T. (2007). Improved estimators of hedonic housing price models. *Journal of Real Estate Research*, *29*(3), 267-302.

Bao, H. X., & 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.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199-231.

Bitter, C., Mulligan, G. F., & Dall'erba, S. (2007). Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems*, *9*(1), 7-27.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 211-252.

Can, A. and Megbolugbe, I. (1997). Spatial Dependence and House Price Index Construction. *Journal of Real Estate Finance and Economics*. 14(1/2), 203-222.

Carroll, T. M., Clauretie, T. M., and Jensen, J. (1996). Living Next to Godliness:Residential Property Values and Churches. *Journal of Real Estate Finance and Economics*. 12(3), 319-330.

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

Costello, G., & Watkins, C. (2002). Towards a system of local house price indices. *Housing Studies*, *17*(6), 857-873.

Clapp, J. M. (2004). A semiparametric method for estimating local house price indices. *Real Estate Economics*, *32*(1), 127-160.

Clapham, E., Englund, P., Quigley, J. M., & Redfearn, C. L. (2006). Revisiting the past and settling the score: index revision for house price derivatives. *Real Estate Economics*, *34*(2), 275-302.

Coulson, T., Ezard, T. H. G., Pelletier, F., Tavecchia, G., Stenseth, N. C., Childs, D. Z., ... & Crawley, M. J. (2008). Estimating the functional form for the density dependence from life history data. *Ecology*, *89*(6), 1661-1674.

de Haan, J., & Diewert, W. E. (2011). Handbook on residential property price indexes. *Luxembourg: Eurostat*.

Fletcher, P., Gallimore, P., and Mangan, J. (2000). The Modelling of Housing Submarkets. *Journal of Property Management*. 18(5), 366-374.

Goh, Y. M., Costello, G., & Schwann, G. (2012). Accuracy and robustness of house price index methods. *Housing Studies*, *27*(5), 643-666.

Goodman, A. C. (1978). Hedonic prices, price indices and housing markets.*Journal of Urban Economics*, *5*(4), 471-484.

Hill, R. J. (2013). Hedonic price indexes for residential housing: A survey, evaluation and taxonomy. *Journal of Economic Surveys*, *27*(5), 879-914.

Hannonen, M. (2006). An analysis of trends and cycles of land prices using wavelet transforms. *International Journal of Strategic Property Management*,*10*(1), 1-21.

Halvorsen, R., & Pollakowski, H. O. (1981). Choice of functional form for hedonic price equations. *Journal of Urban Economics*, *10*(1), 37-49.

Jones, J. and Casetti, E. (1992). *Applications of the Expansion Method*. London: Routledge.

Knight, J. R., Hill, R. C., & Sirmans, C. F. (1992). Biased prediction of housing values. *Real Estate Economics*, *20*(3), 427-456.

Knight, J. R., Carter Hill, R., & Sirmans, C. F. (1993). Stein rule estimation in real estate appraisal. *Appraisal Journal*, *61*, 539-539.

Malaysian Government (1996), Seventh Malaysia Plan, Percetakan Nasional Berhad, Kuala Lumpur.

Malaysian Government (2001), Eight Malaysia Plan, Percetakan Nasional Berhad, Kuala Lumpur.

Meese, R., & Wallace, N. (1991). Nonparametric estimation of dynamic hedonic price models and the construction of residential housing price indices. *Real Estate Economics*, *19*(3), 308-332.

Maurer, R., Pitzer, M., & Sebastian, S. (2004). Hedonic price indices for the Paris housing market. *Allgemeines Statistisches Archiv*, *88*(3), 303-326.

Namba, A., & Ohtani, K. Small Sample Properties Of A Pre-Test Stein-Rule Estimator For Each Individual Re-Gression Coefficient Under An Alternative Null Hypothesis In The Pre-Test. *Kobe University Economic Review*, *58*, 1-9.

Nappi-Choulet, I., & Maury, T. P. (2009). A Spatial and Temporal Autoregressive Local Estimation for the Paris Housing Market.

Pace, R. K. (1993). Nonparametric methods with applications to hedonic models. *The Journal of Real Estate Finance and Economics*, *7*(3), 185-204.

Pace, R. K., Barry, R., Clapp, J. M., and Rodriguez M. (1998). Spatiotemporal Autoregressive Models of Neighborhood Effects. *Journal of Real Estate Finance and Economics*. 17(1), 15-33.

Pavlov, A. D. (2000). Space-varying regression coefficients: a semi-parametric approach applied to real estate markets. *Real Estate Economics*, *28*(2), 249-283.

Pakes, A. (2003), “A Reconsideration of Hedonic Price Indexes with an Application to PCs”, *American Economic Review* 93(5), 1576-93.

Rodriguez, M. and Sirmans, C. F. (1994). Quantifying the Value of a View in Single- Family Housing Markets. *Appraisal Journal*. 62(4), 600-603.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *The journal of political economy*, 34-55.

Schäfer, J., & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, *4*(1).

Song, H. S., & Wilhelmsson, M. (2010). Improved price index for condominiums. *Journal of Property Research*, *27*(1), 39-60.

Widlak, M., & Tomczyk, E. (2010). Measuring price dynamics: evidence from the Warsaw housing market. *Journal of European Real Estate Research*, *3*(3), 203-227.

Panik, M.(2009): Regression Modeling-Methods, Theory, and Computation with SAS. *Statistical Papers*, *53*(3), 803-804.

Xu, T. (2008). Heterogeneity in Housing Attribute Prices: A Study of the Interaction Behaviour between Property Specifics, Location Coordinates and Buyers' Characteristics. *International Journal of Housing Markets and Analysis.* 1(2),166-181.

Email Contact:chengwesley0314@gmail.com