The Construction of Real Estate Price Index: Modelling by Incorporating Spatial Elements

Mohd Lizam¹, Norshazwani Afiqah Rosmera², Abdul Jalil Omar³

Faculty of Technology Management and Business ¹
Universiti Tun Hussein Onn Malaysia
Batu Pahat, Johor
lizam@uthm.edu.my

Faculty of Technology Management and Business ²
Universiti Tun Hussein Onn Malaysia
Batu Pahat, Johor
hp110062@siswa.uthm.edu.my

Faculty of Technology Management and Business ³
Universiti Tun Hussein Onn Malaysia
Batu Pahat, Johor
jalil@uthm.edu.my

Abstract — Housing plays an important role in the economics of developed countries. It is a major source of value and of risk to household and banking system sector. Of these reason, most countries develop a House Price Index (HPI) in order for them to monitor the house price changes from a period to another. In Malaysia, a HPI known as Malaysian House Price Index (MHPI) is constructed. The MHPI is the only available HPI which makes it as the only reference in monitoring the house price changes. The purpose of this study is to construct a HPI with the focus to incorporate spatial elements. Based on the hedonic price modelling, three Hypothetical HPI (HHPI) are constructed. Each HHPI composed of different spatial elements used as variable. Spatial elements included in this study are sub-districts dummy, distance of house to city centre and coordinate (x,y) of house. Results show that the inclusion of house coordinate helps to explain the exact location of houses and smoothen the movement of HPI. This study revealed the significance of considering house coordinate in estimating the house prices, hence producing an accurate HPI.

Keywords — House price index; hedonic price modelling; spatial elements

I. Introduction

House prices play a critical role in understanding the dynamic of housing market. The importance of house prices to understand the complexity of housing market lead to the construction of House Price Index (HPI). To date, the HPI has become an important tool to many parties involved in real estate market. These include investors, financial institution, researchers and developers.

Generally, the main purpose of HPI is to monitor the changes of house prices from period to another. Netzell (2010) supported the use of HPI is to monitor the real estate cycle and relationship between real estate markets. Different parties may use HPI for different purposes. Parties such as investors, financial institution, researchers, policy maker and developers depend on price index for a specific purpose (Gourieroux & Laferriere, 2009).

Policy maker uses HPI prior to formulation of economics as housing market contributes to GDP significantly. They rely on the property price signals to do a decision making. In the perspective of investors, price index is used to benchmark and monitor the equity investment. Other than that, Longford (2009) posited that HPI play a significant role in individual decision whether to buy or to sell a property.

Due to the significance of HPI, most developed countries produced HPIs. Such countries include UK and US in which the price indices have been established over 40 years. In the UK, the application of HPI is seen as early as 1973 (Lim & Pavlou, 2007). Examples of HPIs available in the UK are Land Registry HPI, Halifax HPI, Nationwide HPI and IPD Property Index.

In the US, HPI is constructed due to the needs in monitoring real estate price changes. For instance, US Federal Housing Finance Agency introduced HPI to measure the movement of house prices for single family. Other than that, Freddie Mac HPI (FMHPI) was introduced since 1975 in order to measure the house price inflation.

Kamaruddin et al. (2008) noted in Malaysia, the production of HPI began in 1997 though the attempt to produce it has started since 1993. The HPI was introduced by the Valuation and Property Services Department (VPSD) and it is known as the Malaysian House Price Index (MHPI).

With the purpose of monitoring the real estate price changes from one period to another and assists in formulation of economic policy, the MHPI comprise price indices for 13 states and 2 federal territories (Kamaruddin et al., 2008 and Tan, 2011). The MHPI is constructed based on the hedonic method; a widely used method in US and UK HPI construction.
Since housing is important in the economy, HPI should possess some quality so that it could provide an accurate house price movement (Bourassa, Hoesli & Sun, 2006). Many methodologies related to HPI construction were introduced. Such methods include the hedonic method, repeat-sales method and median price method. Among these methods, hedonic method has drawn a particular attention and widely researched.

The hedonic method has been pioneered by Griliches and in 1974 it was formalised by Rosen. Rosen (1974) noted that hedonic method is based on hypothesis of; products or goods are valued according to their respective characteristics. For housing, hedonic method involves regressing sale price on a vector of house characteristics (Dorsey et al., 2010). The hedonic method can be categorized into two approaches (Bourassa et al., 2006). These are time-dummy method and imputation method.

Although hedonic method is a widely used method, it may lead to spatial effects problems (Long, Paez & Farber, 2007). Spatial effects exist in the property data when one property influenced other in term of the market price, resulting to autocorrelation in the outcome model. Anselin (1999) explained spatial autocorrelation as the coincidence of value similarity with locational similarity.

In the context of Malaysian housing market, houses located in same housing scheme tend to have similar values as they exhibit similar characteristics. As the distance between houses increase, spatial autocorrelation between them will decrease. This situation followed Tobler’s first law of geography; everything is related to everything else, but near things are more related than distant things (Tobler, 1970).

Location of houses plays a significant role in contributing to house prices. Kiel and Zabel (2008) noted house characteristics are spatially related in the form of locational hierarchy. Therefore, in order to ensure the accuracy of constructed HPI, proper spatial data must be considered. Distance, location and topology arrangement are example of spatial data (Gerkman, 2010).

Recently, studies related to property price modelling used distance to the nearest Central Business District (CBD), distance to the sub-markets and distance to the nearest public facilities to explain location of property (Gallimore, Flechter & Carter, 1996). On the other hand, studies done by Gelfand et al. (2004) and Ting (2008) have incorporated the geocoded data; coordinate \((x,y)\) of property and it shows that the variable is significant in house price modelling.

## II. Methodology

### A. Data

This study is based on secondary data obtained from NAPIC. It consists of transaction data for residential property covering Kuala Lumpur and Klang Valley area. A total of 5,365 data of double-storey terraced house is made available for this study. The original data is compiled in annual files; begin from year 2005 to year 2012. Transaction information such as the transaction price, date of transaction, house area, year built and owner specific characteristics are available.

The original transaction data has been reduced through the data cleaning process stages. During this stage, invalid data points are eliminated. This includes data with missing values, inconsistent data and data that did not meet the mathematical control limit. As a result, only 2,000 data are available for the analysis. Besides, only seven property information that covers the physical and locational characteristics of property can be used in the HPI modelling.

The HPI is modelled on a quarterly basis begin with Quarter1: 2005 to Quarter2: 2012. Details of property information used are listed in Table I.

### TABLE I. DETAILS OF PROPERTY INFORMATION

<table>
<thead>
<tr>
<th>Property Information</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>LotArea</td>
<td>Size of land of the observed property</td>
</tr>
<tr>
<td>BuildingArea</td>
<td>Overall building size of property</td>
</tr>
<tr>
<td>Age</td>
<td>Age of property measured by differenting the transaction year and completion year of property</td>
</tr>
<tr>
<td>Bedr</td>
<td>Total number of bedrooms available in observed property</td>
</tr>
<tr>
<td>Distance</td>
<td>Distance of property to the nearest city centre</td>
</tr>
<tr>
<td>Sub-district</td>
<td>Sub-districts of the observed property. Sub-districts available in this study are Kuala Lumpur, Batu and Petaling.</td>
</tr>
<tr>
<td>Coordinate</td>
<td>Describe the exact location ((x,y)) of property. It is based on the Universal Transverse Mercator (UTM) system; Easting and Northing</td>
</tr>
<tr>
<td>Time</td>
<td>Transaction date of respected property. Describe in quarterly form.</td>
</tr>
</tbody>
</table>

Table I defined the property information used as variables in the HPI modelling. Basically, it covers all house characteristics; physical, locational and time. Osland (2010) noted that there is no specific property characteristic need to be included in the HPI modelling.

### B. The HPI Modelling

Based on hedonic method formalised by Rosen in 1974, multiple regression is employed to model the HPI. Fundamentally, multiple regression will show the relationship between dependent variable and its explanatory variables. In this study, the dependent variable is the transaction price of property whilst the explanatory variables are the house characteristics as showed in Table I above. The relationship is given as:

\[
Price = f(\text{physical, locational, time-dummy})
\]  

Equation (1) shows the relationship between the transaction price and house characteristics which represent the explanatory variables. The multiple regression equation for the above relationship is given as:

\[
y_i = \alpha + \sum_{k=1}^{p} \beta_{ki} x_{ki} + \beta_{x2} x_{2i} + \beta_{x3} x_{3i} + \cdots + \beta_{xu} x_{ui} + e_i
\]  

(2)
Where \( i = 1, 2, 3, \ldots, n \) (\( n \) = number of observations), \( y_i \) is the transaction price for \( i \)th house, \( \beta_{ki} \), where \( k = 1, 2, \ldots, k \) represent the previous house characteristics, \( \beta, \beta_2, \ldots, \beta_k \) is the determined coefficient for house characteristics, \( x_{it}, t = 2, \ldots, t \) is the time-dummy variable where a dichotomous value of 1 is used when the house is sold in period \( t \). Otherwise, zero value is used. Price index for period \( t \) is obtained by computing the anti-log \( \beta_t \).

Equation (2) is used to construct three Hypothetical House Price Indices (HHPIs). Each model composed of different locational characteristics used as the explanatory variables. For the first HHPI, two commonly used locational attributes are included. For these are distance to city centre and sub-district dummy of property. An additional of property coordinate as explanatory variable can be seen in the second HHPI. Finally, in the third HHPI, only coordinate of property is used as the locational attribute.

The constructed HHPIs are given as:

**HHPI Model I:**

\[
\text{logPrice}_i = \alpha + \beta_1 \text{logLotArea}_i + \beta_2 \text{logBuildingArea}_i + \beta_3 \text{Age}_i + \beta_4 \text{Age}^2_i + \beta_5 \text{Bedr}_i + \sum_{k=1}^{30} \beta_{ki} \text{subdistrict}_i + \beta_6 \text{Distance}_i + \sum_{k=1}^{30} \beta_{ki} \text{Time}_i + \epsilon_i
\]

**HHPI Model II:**

\[
\text{logPrice}_i = \alpha + \beta_1 \text{logLotArea}_i + \beta_2 \text{logBuildingArea}_i + \beta_3 \text{Age}_i + \beta_4 \text{Age}^2_i + \beta_5 \text{Bedr}_i + \sum_{k=1}^{30} \beta_{ki} \text{subdistrict}_i + \beta_6 \text{Distance}_i + \beta_7 X_{\text{Eastin}} + \beta_8 Y_{\text{Northing}} + \sum_{k=1}^{30} \beta_{ki} \text{Time}_i + \epsilon_i
\]

**HHPI Model III**

\[
\text{logPrice}_i = \alpha + \beta_1 \text{logLotArea}_i + \beta_2 \text{logBuildingArea}_i + \beta_3 \text{Age}_i + \beta_4 \text{Age}^2_i + \beta_5 \text{Bedr}_i + \beta_6 X_{\text{Eastin}} + \beta_7 Y_{\text{Northing}} + \sum_{k=1}^{30} \beta_{ki} \text{Time}_i + \epsilon_i
\]

### III. Result and Discussion

#### A. The Descriptive Analysis

This section begins with the discussion of descriptive statistics of all attributes used in the models. Table II listed out the overall descriptive statistics of attributes used in modelling the HPI.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transaction price (RM)</td>
<td>568610</td>
<td>186099</td>
<td>290000</td>
<td>1100000</td>
</tr>
<tr>
<td>Lot Area (sq.m)</td>
<td>180.61</td>
<td>52.98</td>
<td>130</td>
<td>497</td>
</tr>
<tr>
<td>Building Area (sq.m)</td>
<td>180.00</td>
<td>29.84</td>
<td>148</td>
<td>400</td>
</tr>
<tr>
<td>Age (year)</td>
<td>19</td>
<td>11.93</td>
<td>1</td>
<td>44</td>
</tr>
</tbody>
</table>

The lowest house price as shown in Table II is RM290,000 and highest price is RM1.1 million. The variations of house prices in the observed area are quite large. This may due to the house price changes between period of 2005 and 2013. For lot area, the minimum size is 130 sq.m whilst the maximum size is 497 sq.m. The minimum size of lot area normally represents the intermediate unit of house. On the other hand, larger size represents corner lot house.

In term of building area attribute, it comes with minimum size of 148 sq.m and maximum size of 400 sq.m. Large building area indicates that the house is a corner unit which the homeowner may extend the building on the extra land.

Building age is included in the model to see the depreciation of the building. In this study, minimum building age is one year and maximum is 44 years. Normally, the older building will have lower price as compared to the new building as it may have dilapidated.

Also noted from the Table II, there is a small variation of bedroom number in the observed property. The minimum number of bedroom is three and the maximum number is 6. The mean value for bedroom number is four which indicates most of the houses are designed with four bedrooms.

Distance of house to nearest city centre represents the locational characteristic of houses. The minimum distance is 3 km and maximum is 23 km. The farther away the houses from city centre, the lower the house price. Most houses are located 13 km away from the city centre. This is shown from the mean value.

#### B. The Empirical Analysis

Based on equation (3), (4) and (5), hedonic regression is conducted. The analysis is run based on the Ordinary Least Square (OLS) estimator. Table III shows the empirical result for each model:

<table>
<thead>
<tr>
<th>HHPI Models</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R^2 )</td>
<td>0.5938</td>
<td>0.8171</td>
<td>0.6840</td>
</tr>
<tr>
<td>Estimated error variance</td>
<td>0.0552</td>
<td>0.8614</td>
<td>0.0594</td>
</tr>
<tr>
<td>Breusch-Pagan/Cook Weisberg Test</td>
<td>0.0515</td>
<td>0.9248</td>
<td>0.1379</td>
</tr>
<tr>
<td>No. of observations</td>
<td>1214</td>
<td>1394</td>
<td>1300</td>
</tr>
</tbody>
</table>

| \( \text{logLotArea} \) | (0.000) | (0.000) | (0.000) |
| \( \text{logBuildingArea} \) | (0.000) | (0.000) | (0.000) |
| \( \text{Age} \) | (0.000) | (0.000) | (0.000) |
| \( \text{Age}^2 \) | (0.000) | (0.000) | (0.000) |
| \( \text{Bedr} \) | (0.001) | (0.001) | (0.000) |
| \( \text{Distance} \) | (0.00) | (0.00) | (0.00) |
| \( X_{\text{Eastin}} \) | -0.0748 | -0.0465 | - |
| \( Y_{\text{Northing}} \) | -0.0442 | 0.0388 | |

TABLE III. EMPIRICAL ANALYSIS OF THE CONSTRUCTED HHPI MODELS
The $R^2$ value for HHPI Model I, Model II and Model III are 0.5938, 0.8171 and 0.6840 respectively. The highest $R^2$ value is represented by HHPI Model II. From the result, it indicates that 82% of the house prices are contributed from house attributes used as variables. The additional of coordinates ($X_{\text{East}}$,$Y_{\text{North}}$) in HHPI Model II helps to increase the $R^2$ value.

The hedonic regression result also shows that all $p$-value are less than 5%. It indicates that all explanatory variables included in the models are significant and contribute to the property prices. The coefficient value for each variables show similar results produced by past studies in the same area.

Besides, residuals for all HHPI models exhibit homoskedasticity. This can be seen from results produced by the Breusch-Pagan/Cook-Weisberg test. Results for all models are higher than significance level of 0.05. Diagnostic tests on zero correlation assumption are also conducted. Results show that residuals and explanatory variables for all models are not perfectly correlated.

A normality assumptions test is also conducted. Results show that all HHPI models are normally distributed. Therefore, from all results produced by the diagnostic test and hedonic regression, all constructed HHPI models are valid and fit.

C. The Hypothetical HPI

The construction of HHPI is based on a quarterly basis. A total of 30 quarters are available. The index begins from Quarter 1: 2005 to Quarter 2: 2012. The base period for each HPI is at first quarter 2005. The house price indices are constructed using coefficient value of time dummies produced by the hedonic regression. The HHPI are created by taking the antilog of the estimated time-dummy coefficient for each time period.

![HHPI models as compared to the MHPI](image)

Fig. 1 compares the movement of house prices between Q1:2005 and Q2:2012 for each HHPI models and MHPI for Kuala Lumpur. The MHPI, HHPI Model I, HHPI Model II and HHPI Model III are represented by the blue line, red line, green line and purple line respectively. The patterns of all indices are pointed to upwards trending.

Generally, there are significant differences in the trending patterns produced by the HHPI models as compared with the MHPI. HHPI Model I and Model II diverge quite large from MHPI. The diverging patterns become clear started from Q4: 2008. On contrary, HHPI Model III shows the most similar pattern with MHPI. The index value produced by the model did not contradict too large from MHPI.

From the HPI pattern, it shows that by including all locational attributes as presented by HHPI Model III, it helps to smooth the index. The coordinate $(x,y)$ of houses helps to improve the constructed HPI.

The dissimilarity between the constructed HHPI models and MHPI might arise due to the difference data set used in the construction of HPI. The MHPI covers all area in Kuala Lumpur whereas the constructed HHPI only cover on three sub-districts; sub-districts Batu, Kuala Lumpur and Petaling.

Besides, the difference in the price indices may arise due to the house characteristics used as the explanatory variables. Some variables such as house type, building quality and tenure type used to construct MHPI are not available in the data set used in this study. The shortage of house information is due to the limitation of data provided by NAPIC.

IV. Conclusion

The concern of this study is to incorporate spatial elements in constructing HPI. Three different models of HHPI are constructed which each model composed of different locational attributes. Results produced from the hedonic regression show that the inclusion of coordinate $(x,y)$ helps to improve the accuracy of HPI.

Besides, the significance of coordinate $(x,y)$ in modelling the HPI can be seen in term of $R^2$ value produced from the analysis. The inclusion of coordinate $(x,y)$ has increased the $R^2$ value. This indicates that coordinate $(x,y)$ contribute to the house prices.

Results produced from this study have shown the significance of considering house coordinate in property price modelling as it helps to improve the accuracy level of HPI. There is a huge potential to include spatial elements in HPI modelling. It is hoped that data provider such as VPSD and NAPIC can provide the property coordinate so that the accuracy of HPI could be improved.

Acknowledgment

The authors would like to thank the Ministry of High Education (MOHE) and Office of Research, Innovation, Commercialization and Consultancy (ORICC) University Tun Hussein Onn Malaysia (UTHM) for supporting this research.
under the Fundamental Research Grant Schemes (FRGS) (Vote No. 1067). Deep appreciation is also extended for the organization that was willing to provide the necessary data needed for the research. The utmost appreciation goes to National Property Information Centre (NAPIC) for their willingness in providing and sharing the transaction data needed.

References


