Analysis of housing prices in Petaling district, Malaysia using functional relationship model

Yun Fah Chang and Wei Cheng Choong
Department of Mathematical and Actuarial Sciences, Universiti Tunku Abdul Rahman, Kajang, Malaysia
Sing Yan Looi
Department of Physical and Mathematical Science, Universiti Tunku Abdul Rahman, Kampar, Malaysia
Wei Yeing Pan
Department of Mathematical and Actuarial Sciences, Universiti Tunku Abdul Rahman, Kajang, Malaysia, and
Hong Lip Goh
Department of Building and Property Management, Universiti Tunku Abdul Rahman, Kajang, Malaysia

Received 13 December 2018
Revised 16 June 2019
Accepted 18 June 2019

Abstract

Purpose – The purpose of this paper is to analyse and predict the housing prices in Petaling district, Malaysia and its six sub-regions with a set of housing attributes using functional relationship model.

Design/methodology/approach – A new multiple unreplicated linear functional relationship model with both the response and explanatory variables are subject to errors is proposed. A total of 41,750 housing transacted records from November 2008 to February 2016 were used in this study. These data were divided into 70% training and 30% testing sets for each of the selected sub-regions. Individual housing price was regressed on nine housing attributes.

Findings – The results showed the proposed model has better fitting ability and prediction accuracy as compared to the hedonic model or multiple linear regression. The proposed model achieved at least 20% and 40% of predictions that have less than 5% and 10% deviations from the actual transacted housing prices, respectively. House buyers in these sub-regions showed similar preferences on most of the housing attributes, except for residents in Shah Alam who preferred to stay far away from shopping malls, and leasehold houses in Sri Kembangan are more valuable. From the h-nearest houses indicator, it is concluded that the housing market in Sungai Buloh is the most volatile in Petaling District.

Research limitations/implications – As the data used are the actual housing transaction records in Petaling District, it represents only a segment of Malaysian urban population. The result will not be generalized to the entire Malaysian population.

Practical implications – This study is expected to provide insights to policymakers, property developers and investors to understand the volatility of the housing market and the influence of determinants in different sub-regions. The potential house buyers could also use the model to determine if a house is overpriced.

The authors would like to thank the Jordan Lee and Jaafar (S) Sdn. Bhd. who provided the housing data for this study.

International Journal of Housing Markets and Analysis
© Emerald Publishing Limited
1753-8270
DOI 10.1108/IJHMA-12-2018-0099

Originality/value – This study introduces measurement errors into the housing attributes to provide a more reliable analysis tool for the housing market. This study is the first housing research in Malaysia that used a large number of actual housing transaction records. Previous studies relied on small survey samples.

Keywords Malaysia, Housing prices, Measurement errors, Functional relationship model, Housing attributes, Petaling district

Paper type Research paper

1. Introduction

The housing sector is very much synchronized with a nation's economic health. The fluctuation in demand would affect growth in other economic sectors. Therefore, it is essential that these price model are as accurate as possible as housing price prediction are critical for the investors, regulators and household to understand the property market dynamics, as well as assisting them in decision-making (Clapp and Giaccotto, 1995).

The housing market is different from ordinary consumption goods as it harbour the attributes of locational fixity, relatively long durability and heterogeneity. The housing market also susceptible to macro factors such as demographic qualities and economic development (Hamid and bin Mar, 2006). Various efforts has taken to model housing prices, for example, monecentric model which based on the assumption that the housing price is a proximity of localized employment and a reflection of relative spending power in relates to surrounding areas (Chin and Chau, 2003). In the other hand, the micro-perspective of housing price can be examined by using hedonic pricing method (HPM) approach which postulates that commodities are characterized by their attributes; hence, the total value of the commodity is a summation of values from its constituent attributes. HPM was widely used in estimating housing price, but not without controversies. Previous literature mostly proposed that housing price partly determined by locational and physical characteristics. This includes distance to amenities or services, travel time or travel cost (Herath and Maier, 2010) and the results obtained from HPM are likely to be biased due to several assumptions has to be made when modeling the indexes which includes constant values for some of the attributes.

Most of the previous studies considered housing attributes as fixed values when modeling the housing prices. However, some attributes such as distance to the nearby amenities and housing age may subject to error. The distances between houses to amenities might be slightly different and housing age might subject to refurbishment and renovation. This may result in developing a housing price model that cannot truly explain the actual situation of the housing market behavior. By limiting the housing attributes as fixed value also provide challenges in explaining situations where houses with similar housing attributes may be sold at different prices and vice versa. Perhaps, a study on the interrelationship between housing prices and housing attributes might provide more insight on explaining such phenomena.

This study proposes a new functional model to investigate the relationship between housing prices and a set of micro-perspective determinants. The proposed model is taking in consideration of errors associated with variables while using HPM model and able to provide greater efficiency in predicting housing price when compared with multiple regression model (MR). The proposed model was aimed to investigate the influence of housing attributes on property price in six sub-regions of Petaling District, Malaysia. The use of a functional relationship model allows the interrelationship between dependent and independent variables to be studied. In this study, house price predictions are highly influenced by extreme values, thus the idea of using the average house price of a certain number of nearest houses is proposed to diminish or eliminate the effect of extreme prices and hence increase prediction accuracy.

2. Literature review

2.1 Hedonic pricing method and international housing

Many other empirical types of research that used the hedonic models have been conducted in both developed and developing nations. Among others are Clark and Herrin (2000), Clauretie and Neill (2000), Gabriel (1984), Ketkar (1992), Megbolugbe (1989), So *et al.* (1996), Ottensmann *et al.* (2008) and Tyrvainen (1997).

Monson (2009) used hedonic models to study the condominium price in South Boston, the office building in Peoria, IL, and multi-family condominium units in Reston, VA in the USA. Among 22 attributes considered, Monson (2009) concluded that attached garage, swimming pool, private outdoor space, security systems, and extra storage space are statistically significant variables contributed to transaction condominium price in South Boston. In the study for the office building in Peoria, 10 variables over 280 office properties were analyzed and the significant variables are total building square footage, real commercial property index (CPPI), green technology, year renovated and class of the building. There were 154 multi-family condominium units with 5 variables selected in Reston. However, the prediction of house price using the hedonic models is moderately accurate with an average difference of 10 per cent from the actual transaction house price.

Efforts to improve on prediction efficiency of HPM are not new. Ismail and MacGregor (2005) studied the housing markets in Glasgow, Scotland combining hedonic model with the aid of Geographical Information System (GIS) and spatial statistics. The developed model regressed 2,715 housing prices against 61 independent variables mainly from the microeconomic perspective for the year 2002. They claimed that multicollinearity, heteroscedasticity and spatial autocorrelation are the main sources of problems in a housing price hedonic analysis. To improve the imperfection, GIS has been used in this model as it involved spatial statistics which can detect positive spatial autocorrelation easily. As a result, the spatial hedonic has improved the adjusted R^2 by 3.9 per cent point from 75.8 per cent under the ordinary least square model.

A new hedonic model consists of the location, and individual fixed effects were proposed by Jiang *et al.* (2014). This new hybrid approach is less prone to specification errors and greater computational efficiency. Jiang *et al.* (2014) fit the model to private single-sale and repeat-sale properties in Singapore between 1995 and 2014. The hybrid hedonic model slightly outperforms the Case-Shiller index (Case and Shiller, 1987, 1989) in predicting the price of single-sales homes out-of-sample, but less accurate for repeat-sales homes out-of-sample case.

Adyan *et al.* (2017) studied the house price of nine houses located in Malang, East Java, Indonesia with 15 micro- (home id, street name, building area and et cetera) and two macro-economic (value of selling tax object building and land prices) determinants. In their research, they had adopted linear regression and particle swarm optimization methods to perform house price prediction of seven different areas – Kelurahan Karang Besuki, Tunggulwulung, Lowokwaru, Puncak Trikora, Sumbersari, Dinoyo and Manggar, with seven models each of which represented an area. Among the models developed, model that represented Kelurahan Karang Besuki has shown the best prediction with a root mean square error of IDR 14.186. However, the reliability of the finding is to be questioned as the sample size is too small.

Neelam and Kiran (2018) on the other hand studied the house price of 3,000 houses located in Ames, IA with 37 microeconomic housing attributes obtained from the Kaggle database using regression analysis. In their research, they had estimated the house prices using linear regression, support vector regression, lasso regression, and decision tree, and the performance and prediction accuracy of those models were compared and evaluated using mean square error (MSE), R^2 value and et cetera. Among those models, the decision tree has shown the best performance with the highest accuracy where $R^2 = 0.99$ and MSE = 47184.93 compared to linear regression with $R^2 = 0.987$ and MSE = 79604145.

2.2 The studies of Malaysian housing market

The MHPI was first introduced in 1997 by the National Property Information Centre (NAPIC). The house price index is a transaction-based index which was computed using Laspeyres weighted formula (Francis, 2004) with 1990 as the base year and it measures the change in prices which has been paid for an "average" house using linear regression analysis or known as the hedonic model. The change in prices was estimated by pricing the "average" house at the current year and comparing to the base year with the intention to know how much of the cost of housing has changed between these two years by assuming the house buyers maintained their standard of living as of the base year. The mentioned "average" house was priced according to a set of fixed characteristics which comprised of variables mainly from the micro perspectives. The MHPI was then revised and re-based to the year 2000 to reflect the changes in buyers' preference which is to show the new trend in the housing market. Under the revised MHPI, two statistical techniques namely principal component analysis and two-step cluster analysis were used to analyze the characteristics or more precisely the housing neighborhoods which included physical, environmental, social and economic characteristics (NAPIC, 2015).

The macro perspective of Malaysian housing price prediction was covered by several researchers. Prior to 2000, a research done by Tan (1999) studied the effects of economic and financial factors on the housing prices using multiple regression model. The study showed that total loans to housing, per capita income, unemployment rate, and Kuala Lumpur Stock Exchange (KLSE) composite index are significant determinants of MHPI. Step-wise model selection was adopted to eliminate the effects of multicollinearity between explanatory variables. However, the study failed to provide a good justification for its findings of the positive relationship between the unemployment rate and MHPI. The causal relationship of property shares and stocks and housing price are again supported by Lee and Ting (2011), indicating future policy on listed property companies would likely influence the development of housing prices. Fundamental determinants such as income and population growth again showed as strong determinants for supply and demand for housing market in Wong *et al.* (2019).

Ong (2013) adopted a regression analysis to study six macro-determinants of MHPI, which are inflation rate, GDP, population, cost of construction, interest rate, and real property gain tax (RPGT). Ong (2013) concluded that MHPI is significantly correlated to GDP, population and RPGT. Study by Shiau *et al.* (2018) also echoed that GDP do contributed to the fluctuation of housing prices, while availability of money due to national monetary policy would stimulate housing demand due to economic prospects. Other local studies on macro factors include comprehensive study conducted by Central Bank of Malaysia (2013) has looked into the effects of thirteen macroeconomic-factors, financial-factors, and government regulations and policies on housing prices using a multivariate regression model. In this model, economic growth, change in demography and inflation rate are dominating the change in housing prices. Contrarily, the study of Central Bank of Malaysia found that real GDP, inflation rate, base lending rate and RPGT are not significantly contributed to explaining the housing prices in Malaysia. Lim and Chang (2018) highlighted that overall landed properties fared better than multi-storey buildings as preferred residential typology among Malaysian house buyers in urban cities regardless of their socio-economic backgrounds. Study by Lee (2014) also showed that Malaysian residential property able to withstand impact of inflation in the long run, however inconclusive over unexpected inflation. The issue of oversupply might be the possible explanation. Perhaps, further studies are needed.

In 2002, Chau and Chin proposed a hedonic price model by adapting Lancaster's consumer theory and Rosen's model to study the behavior of the housing market in Penang. Chau and Chin (2002) considered only 120 condominium transacted prices in Penang and it was found that actual floor area, floor level, distance from the central business district, proximity to large shopping centre, proximity to the premier school and availability of facilities are key contributors to the housing prices.

Yusof and Ismail (2012) analyzed house price variations of double-storey terrace in Kuala Lumpur using multiple regressions. They concluded that locality is the most influential factor contributed about 80 per cent of the variations of housing prices in Kuala Lumpur city. On the other hand, Yusof and Ismail (2012) compared hedonic models for the double-storey terrace in Penang, Ipoh, Kuala Lumpur, Kuantan and Johor Bahru. Separate models were constructed for years 1990, 1995, 2000, 2002, 2003, and 2004. It was found that the gross domestic product (GDP) growth is the main macro-determinant which affects housing prices in these states with the coefficient of determination, $R^2 > 0.80$. For the micro-perspective, Yusof (2012) claimed that housing prices in different states are affected by different predominant factors. For example, in Kuala Lumpur, the MHPI is well explained by vocational factors while in Johor Bahru, utility-bearing characteristics are found to be the main factor in affecting housing prices.

Norshazwani *et al.* (2013) studied micro-perspective of the housing market in three sub-districts in Kuala Lumpur. They concluded that lot area, building area, house age, and other micro factors are significantly contributing to MHPI. Norshazwani *et al.* (2013) proposed time dummies from the hedonic model to estimate the Hypothetical House Price Index, an alternative to MHPI. However, this hypothetical house price index failed to correlate well with MHPI. A similar study by Kam *et al.* (2016) focused on the double-storey house in Rawang, Selangor, a sub-urban area located 33 km from the north of Kuala Lumpur. Unlike an urban area, terrace house prices in Rawang are affected by built-up area and distance from the shopping centre with the coefficient of determination, $R^2 = 0.668$.

3. Methodology

3.1 Data

This study focuses on terrace houses in Petaling District, the most populated region in Malaysia. Petaling District is in Selangor state and adjacent to Kuala Lumpur city (Figure 1). It is divided into six sub-regions under three administrative zones. Shah Alam and Sungai Buloh are under the governance of Shah Alam City Council, Subang Jaya, Puchong, and Seri Kembangan are under the governance of Subang Jaya City Council, and Petaling Jaya is under the governance of Petaling Jaya City Council. The area was selected due to the common perception where Kuala Lumpur and its surrounding (in this case Petaling District) leads the national residential price cycle and the ripple effect might affect similar projects in greater neighbouring areas as well (Lian and Smyth, 2012).

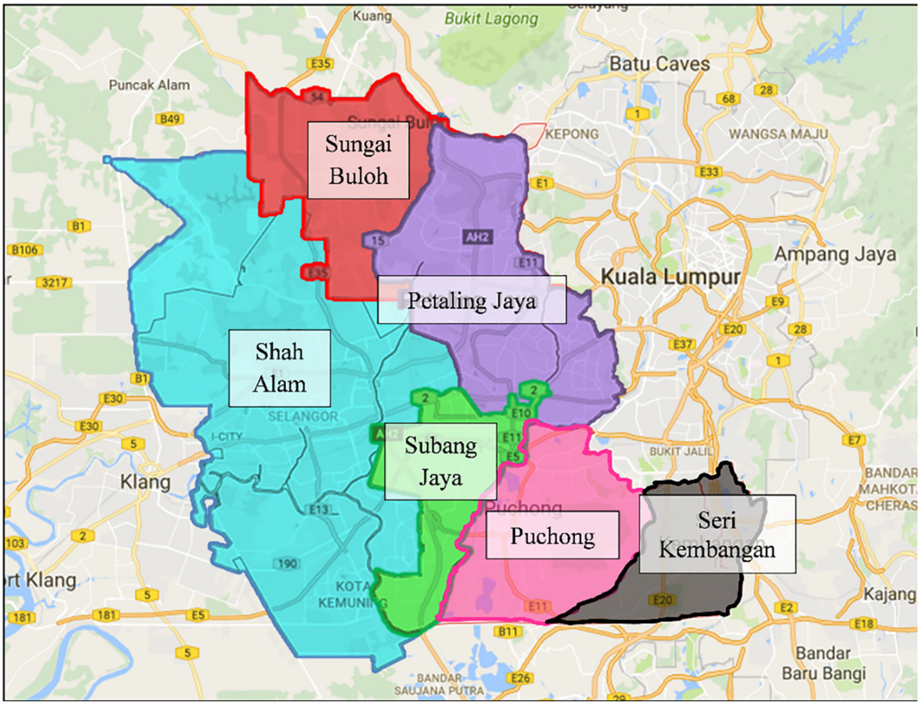


Figure 1.
Administrative map of Petaling District

Source: Google maps

The main data source was collected from Jordan Lee and Jaafar (S) Sdn. Bhd., and it consists of 44,331 terrace house transacted records from November 2008 to February 2016. A total of 1,167 (2.63 per cent) duplicated records and 1,414 (3.19 per cent) records with missing values or non-rectifiable errors were removed. Hence, the remaining 41,750 were used and grouped according to the six sub-regions based on the address given where 9,643 cases from Shah Alam; 9,341 cases from Puchong; 8,741 cases from Petaling Jaya; 7,956 cases from Subang Jaya; 5,477 cases from Seri Kembangan; and 592 cases from Sungai Buloh. These data were randomly divided into a 70 per cent training set and 30 per cent testing set. The training set was used to train the model, and the testing set was used to validate the performance of the trained models.

The relevant information extracted from source data are transacted house prices, lot sizes, tenure types (freehold or leasehold), expiry dates of lease terms, terrace types, number of bedrooms, main building sizes and transaction date. Additional information from Google Maps such as distance to the nearest shopping mall and distance to the nearest supermarket were also included based on the property's address. Table I summarizes the variables used in this study.

Referring to Table II, out of the 41,750 total transacted houses that was analysed, 25.01 per cent sold at prices between RM150,000 to RM300,000, another 24.81 per cent ranged from RM300,000 to RM450,000. Transacted house prices from RM450,000 to RM750,000 account for 16.69 and 11.91 per cent. The remaining are either low cost houses below RM150,000 or luxury houses above RM750,000. For tenure type, 66.33 per cent of the

Variable	Description
y_i	Individual housing price (RM'000)
x_{i1}	Lot size (m^2)
x_{i2}	Tenure type (freehold – 0, leasehold – 1)
x_{i3}	Years to expiry of lease term (assuming 200 years for freehold)
x_{i4}	Terrace type (total number of floors, i.e. 1 for single storey, 2 for double storey)
x_{i5}	Number of bedrooms
x_{i6}	Main building size (m^2)
x_{i7}	Distance to nearest shopping mall (km)
x_{i8}	Distance to nearest supermarket (km)
x_{i9}	Transaction date (in month)

Table I. Descriptions of variables used

Variables	(%)	Variables	Statistic
Transacted house price (RM' 000)		*Lot size (m^2)*	
Below 150	3.90	Mean	160.93
150-300	25.01	Standard deviation	70.72
300-450	24.81	Median	153.00
450-600	16.69	Minimum	43.48
600-750	11.91	Maximum	3617.00
750-900	8.19		
900-1050	3.60	*Expiry of leasehold term (years)*	
1050-1200	1.84	Mean	82.28
1200-1350	1.70	Standard deviation	9.80
At least 1350	2.35	Median	82.78
		Minimum	42.09
Tenure type		Maximum	102.96
Freehold	66.33		
Leasehold	33.67	*Building Size (m^2)*	
		Mean	134.78
Terrace type		Standard deviation	48.14
Single–storey	11.87	Median	131.18
1.5–storey	1.65	Minimum	40.85
Double–storey	81.57	Maximum	571.04
2.5–storey	3.97		
3–storey	0.91	*Distance to nearest shopping mall (km)*	
Others	0.04	Mean	3.76
		Standard deviation	2.60
Number of bedrooms		Median	2.90
1–bedroom	0.01	Minimum	0.014
2–bedrooms	5.40	Maximum	12.3
3–bedrooms	49.86		
4–bedrooms	42.24	*Distance to nearest supermarket (km)*	
5–bedrooms	2.25	Mean	2.44
6–bedrooms	0.23	Standard deviation	1.77
7–bedrooms	0.01	Median	1.90
		Minimum	0.11
		Maximum	9.80

Table II. General information of the transacted houses transacted houses are freehold and the remaining 33.67 per cent are leasehold. Among the leasehold terraces, their average expiry of lease term is 82.28 years with a standard deviation of 9.8. The minimum expiry of lease term is only 42 years while the longest lease term is around 103 years. For terrace type, the double-storey houses comprise of 81.57 per cent of the total transaction record, while the single-storey houses consists of 11.87 per cent. Other terrace types are not popular in Petaling district. On the number of bedrooms, 49.86 and 42.24 per cent of the transacted houses have 3- and 4-bedrooms, respectively.

The average lot size of the transacted houses is 160.93 m^2 with a standard deviation of 70.72. The smallest and the largest lot size being transacted is 43.48 m^2 and 3617 m^2, respectively. When referring to main building size, the average size is 134.78 m^2 and the standard deviation is 48.14, while the minimum building size is 40.85 m^2 and the maximum building size is 571.04 m^2. With regard to the nearest distance between the transacted houses to the amenities, the average nearest distances to shopping mall and supermarket are 3.76 and 2.44 km, respectively. The minimum distance to shopping mall is only 0.014 km, while the longest distance is 12.3 km. For supermarket, the minimum and maximum distance are 0.11 and 9.80 km, respectively.

3.2 The multiple un-replicated linear functional relationship model

In this study, seven models have been developed, where six models were used to study the terraced housing market for each region, while one model was used to study the overall performance of terraced housing market in Petaling District. For each model, 70 per cent of the data were used for training purpose, while the remaining 30 per cent were used for testing purpose. All values of the independent variables were normalized with the intention to compare the coefficients of the housing attributes. The housing prices were regressed against the above-mentioned independent variables using Multiple Un-replicated Linear Functional Relationship (M_PULFR) Model (Choong *et al.*, 2018). It is a special case of MULFR model developed by Chang *et al.* (2010).

Suppose that Y_i is an unobservable housing price and $\boldsymbol{X}_i = (X_{i1}\ X_{i2}\ \cdots\ X_{ip})$ are p unobservable attributes such that:

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} = \alpha + \boldsymbol{X}_i \boldsymbol{\beta},\ \ i = 1, 2, \ldots, n_r \tag{1}$$

where α is an intercept and $\boldsymbol{\beta} = (\beta_1\ \beta_2 \cdots \beta_p)'$ are coefficient of the linear function and n_r is the number of transacted houses in Petaling District or the respective sub-regions.

Consider the two corresponding observed housing prices y_i and observed attributes $\boldsymbol{x}_i = (x_{i1}\ x_{i2} \cdots x_{ip})$ with errors ε_i and $\boldsymbol{\delta}_i = (\delta_{i1}\ \delta_{i2} \cdots \delta_{ip})$ respectively, such that:

$$\left.\begin{aligned} y_i &= Y_i + \varepsilon_i \\ \boldsymbol{x}_i &= \boldsymbol{X}_i + \boldsymbol{\delta}_i \end{aligned}\right\}\quad i = 1, 2, \ldots, n \tag{2}$$

and both error vectors are assumed to be mutually independent and normally distributed with the following properties:

- $E(\varepsilon_i) = 0$ and $E(\boldsymbol{\delta}_i) = \boldsymbol{0}$
- $Cov(\varepsilon_i, \varepsilon_j) = 0$ and $Cov(\boldsymbol{\delta}_i, \boldsymbol{\delta}_j) = \boldsymbol{0}$, $\forall\ i \neq j$
- $Cov(\varepsilon_i, \boldsymbol{\delta}_{ik}) = 0$, $\forall\ i \neq k$, and

- $\varepsilon_i \sim NID(0,\ \omega_{11})$ and $\delta_i \sim NID\ (\mathbf{0},\ \boldsymbol{\omega}_{22})$ where $\omega_{11} = \tau^2$, and $\boldsymbol{\omega}_{22} = \sigma^2 \boldsymbol{I}_p$ then $\boldsymbol{\omega} = \begin{pmatrix} \omega_{11} & \boldsymbol{\omega}_{12} \\ \boldsymbol{\omega}_{21} & \boldsymbol{\omega}_{22} \end{pmatrix}$ where $\boldsymbol{\omega}_{21} = \boldsymbol{\omega}_{12}' = 0$.

The joint density function $(x_{i1}, x_{i2}, \ldots, x_{ip}, y_i)$ or equivalently, (x_i, y_i) is:

$$f(\boldsymbol{x}_i, y_i) = \frac{1}{(2\pi)^{r/2}|\boldsymbol{\omega}|^{1/2}} \exp\left[-\frac{1}{2}\left\{\left[\,(y_i - Y_i) \quad (\boldsymbol{x}_i - \boldsymbol{X}_i)\,\right]\boldsymbol{\omega}^{-1}\begin{pmatrix} y_i - Y_i \\ [\boldsymbol{x}_i - \boldsymbol{X}_i]' \end{pmatrix}\right\}\right]$$

where $r = p + 1$, $E(x_i) = E(\boldsymbol{X}_i + \boldsymbol{\delta}_i) = \boldsymbol{X}_i$, and $E(y_i) = E(Y_i + \varepsilon_i) = Y_i$. For simplicity, let $\tau^2 = \lambda\,\sigma^2$, where λ is a positive constant then the log-likelihood function is:

$$L^* = -\ln K - \frac{n}{2}\ln \lambda - (p+1)n \ln \sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[\frac{1}{\lambda}\left(y_i - \alpha - \boldsymbol{\beta}' \boldsymbol{X}_i'\right)^2 + (\boldsymbol{x}_i - \boldsymbol{X}_i)(\boldsymbol{x}_i - \boldsymbol{X}_i)'\right]$$

where $K = (2\pi)rn/2$, $|\,\omega\,| = |\,\omega_{11}\ \omega_{22}\,| = \lambda\,\sigma^{2(p+1)}$ and $\boldsymbol{\beta}'\boldsymbol{X}_i' = \boldsymbol{X}_i\boldsymbol{\beta}$. Hence, the log-likelihood function with respect to α, β, X_i and σ, and equate them to zero will yield:

$$\hat{\alpha} = \bar{y} - \frac{1}{n}\left(\sum\nolimits_{i=1}^{n} \hat{\boldsymbol{X}}_i\right)\hat{\boldsymbol{\beta}} \tag{3}$$

$$\hat{\boldsymbol{\beta}}' = \left(\sum\nolimits_{i=1}^{n} y_i \hat{\boldsymbol{X}}_i - \hat{\alpha}\sum\nolimits_{i=1}^{n} \hat{\boldsymbol{X}}_i\right)\left(\sum\nolimits_{i=1}^{n} \hat{\boldsymbol{X}}_i' \hat{\boldsymbol{X}}_i\right)^{-1} \tag{4}$$

$$\hat{\boldsymbol{X}}_i = \left[\lambda\,\boldsymbol{x}_i + (y_i - \hat{\alpha})\hat{\boldsymbol{\beta}}'\right]\left(\lambda\,\boldsymbol{I} + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}'\right)^{-1} \tag{5}$$

$$\hat{\sigma}^2 = \frac{1}{(p+1)n}\sum\nolimits_{i=1}^{n}\left[(\boldsymbol{x_i} - \hat{\boldsymbol{X}}_i)(\boldsymbol{x_i} - \hat{\boldsymbol{X}}_i)' + \frac{1}{\boldsymbol{\lambda}}\left(y_i - \hat{\alpha} - \hat{\boldsymbol{X}}_i\hat{\boldsymbol{\beta}}\right)^2\right] \tag{6}$$

To estimate $\hat{\boldsymbol{\alpha}}$, substitute equations (5) into (3) and get:

$$\hat{\alpha} = \bar{y} - \frac{1}{n}\left\{\sum_{i=1}^{n}\left[\lambda\,\boldsymbol{x_i} + (y_i - \hat{\alpha})\hat{\boldsymbol{\beta}}'\right]\left(\lambda\,\boldsymbol{I} + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}'\right)^{-1}\right\}\hat{\boldsymbol{\beta}}$$

$$\hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\beta}}'\left(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}'\right)^{-1}\left(\lambda I + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}'\right) = \bar{\boldsymbol{y}}\hat{\boldsymbol{\beta}}'\left(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}'\right)^{-1}\left(\lambda I + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}'\right) - \frac{1}{n}\sum_{i=1}^{n}\left[\lambda\,\boldsymbol{x}_i + (y_i - \hat{\alpha})\hat{\boldsymbol{\beta}}'\right]$$

$$\hat{\alpha} = \bar{y} - \bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}} \tag{7}$$

To estimate $\hat{\boldsymbol{\beta}}$, substitute equation (5) into equation (4) and rearrange will get:

$$\hat{\boldsymbol{\beta}}' \sum_{i=1}^{n} \left\{ \left[\lambda\, \boldsymbol{x}_i + (y_i - \hat{\alpha})\hat{\boldsymbol{\beta}}' \right] \left(\lambda I + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}' \right)^{-1} \right\}' \left[\lambda\, \boldsymbol{x_i} + (y_i - \hat{\alpha})\hat{\boldsymbol{\beta}} \right] \left(\lambda I + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}' \right)^{-1}$$

$$= \sum_{i=1}^{n} (y_i - \hat{\alpha}) \left[\lambda\, \boldsymbol{x}_i + (y_i - \hat{\alpha})\hat{\boldsymbol{\beta}}' \right] \left(\lambda\, \boldsymbol{I} + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}' \right)^{-1}$$

$$\lambda \sum_{i=1}^{n} \left(\boldsymbol{x}_i \hat{\boldsymbol{\beta}} \right)^2 + \left(\hat{\boldsymbol{\beta}}' \hat{\boldsymbol{\beta}} - \lambda \right) \sum\nolimits_{i=1}^{n} (y_i - \hat{\alpha}) \boldsymbol{x}_i \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}' \hat{\boldsymbol{\beta}} \sum\nolimits_{i=1}^{n} (y_i - \hat{\alpha})^2 = 0 \tag{8}$$

Then, substitute equation (7) into equation (8) and get:

$$\lambda \sum_{i=1}^{n} \left(\boldsymbol{x_i} \hat{\boldsymbol{\beta}} \right)^2 + \left(\hat{\boldsymbol{\beta}}' \hat{\boldsymbol{\beta}} - \lambda \right) \sum_{i=1}^{n} (y_i - \bar{y}) \boldsymbol{x}_i \hat{\boldsymbol{\beta}} - \lambda\, n \left(\bar{\boldsymbol{x}} \hat{\boldsymbol{\beta}} \right)^2 - \hat{\boldsymbol{\beta}}' \hat{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \bar{y})^2 = 0$$

$$\lambda \sum_{i=1}^{n} \left(\sum_{j=1}^{p} x_{ij} \hat{\beta}_j \right)^2 + \left(\sum_{j=1}^{p} \hat{\beta}_j^2 - \lambda \right) \sum_{i=1}^{n} \left[(y_i - \bar{y}) \sum_{j=1}^{p} x_{ij} \hat{\beta}_j \right] - \lambda\, n \left(\sum_{j=1}^{p} \bar{x}_j \hat{\beta}_j \right)^2$$

$$- \sum_{j=1}^{p} \hat{\beta}_j^2 \sum_{i=1}^{n} (y_i - \bar{y})^2 = 0$$

To solve for $\hat{\beta}_k$:

$$\lambda\, \hat{\beta}_k^2 \sum_{i=1}^{n} x_{ik}^2 + \left(\hat{\beta}_k^2 - \lambda \right) \sum_{i=1}^{n} (y_i - \bar{y}) x_{ik} \hat{\beta}_k - \lambda\, n \hat{\beta}_k^2 \bar{x}_k^2 - \hat{\beta}_k^2 \sum_{i=1}^{n} (y_i - \bar{y})^2 = 0$$

$$\hat{\beta}_k = \frac{\left(S_{yy} - \lambda\, S_{x_k x_k} \right) + \sqrt{\left(S_{yy} - \lambda\, S_{x_k x_k} \right)^2 + 4\lambda\, S_{x_k y}^2}}{2 S_{x_k y}} \tag{9}$$

where $S_{x_k x_k} = \sum\nolimits_{i=1}^{n} (x_{ik} - \bar{x}_k)^2$, $S_{yy} = \sum\nolimits_{i=1}^{n} (y_i - \bar{y})^2$, $S_{x_k y} = \sum\nolimits_{i=1}^{n} x_{ik} y_i - n \bar{x}_k \bar{y}$, $\left(\hat{\boldsymbol{\beta}} \hat{\boldsymbol{\beta}}' \right)$ is a reversible and symmetry $p \times p$ matrix and λ is the ratio of the error variances.

Thus, the maximum likelihood estimators are:

$$\hat{\alpha} = \bar{y} - \bar{\boldsymbol{x}} \hat{\boldsymbol{\beta}}$$

$$\hat{\beta}_k = \frac{\left(S_{yy} - \lambda\, S_{x_k x_{\boldsymbol{k}}} \right) + \sqrt{\left(S_{yy} - \lambda\, S_{x_k x_k} \right)^2 + 4\lambda\, S_{x_k y}^2}}{2 S_{x_k \boldsymbol{y}}}$$

$$\hat{\boldsymbol{X}}_i = \left[\lambda\, \boldsymbol{x}_i + (y_i - \hat{\alpha}) \hat{\boldsymbol{\beta}}' \right] \left(\lambda\, \boldsymbol{I} + \hat{\boldsymbol{\beta}} \hat{\boldsymbol{\beta}}' \right)^{-1}$$

Choong *et al.* (2018) has showed that $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are unbiased and consistent estimators, and the estimation of these parameters is not affected by multicollinearity effect which is resulted from the correlation between independent variables. For simplicity, we used $\lambda = 1$ in this study. Chang *et al.* (2010) showed that the estimation of parameters and R^2 still perform well for $\lambda \leq 100$.

Coefficient of determination for M_pULFR model is given as:

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}} \tag{10}$$

where $S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$, and $SS_E = \left\{ \hat{\boldsymbol{\beta}}' \left(\lambda \boldsymbol{I} + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}' \right)^{-1} \left(\lambda \boldsymbol{I} + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}' \right)^{-1} \hat{\boldsymbol{\beta}} + \lambda \left[\hat{\boldsymbol{\beta}}' \left(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}' \right)^{-1} \left(\lambda \boldsymbol{I} + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}' \right)^{-1} \hat{\boldsymbol{\beta}} \right]^2 \right\} \sum_{i=1}^{n} \left(y_i - \hat{\alpha} - \boldsymbol{x}_i \hat{\boldsymbol{\beta}} \right)^2$.

The coefficient of determination measures the total variability of the error variances explained by the model in term of percentage. The R^2 will be used to assess the performance of M_PULFR.

From Choong *et al.* (2018), it is observed that housing price, y_i is used to estimate its respective unobservable attributes, X_i. However, it is always assumed that the price of a house is not known upon prediction. In other words, M_PULFR model requires a reference housing price, $\tilde{y}_m$ when predicting the price of a house. Euclidean distance is used to determine $\tilde{y}_m$:

$$d_i = \sqrt{(\bar{x}_1 - x_{i1})^2 + (\bar{x}_2 - x_{i2})^2 + \cdots + (\bar{x}_p - x_{ip})^2}, i = 1, 2, \ldots, n \tag{11}$$

$$\tilde{y}_m = \frac{1}{h} \sum_{j=1}^{h} \min_{d_i}(y_j) \tag{12}$$

where $\tilde{y}_m$ is the average of historical housing prices of h houses with the most similar housing attributes.

4. Results and discussions

In this section, we will first compare the estimated parameters from M_pULFR and multiple linear regression (MR) models. The attributes that significantly contributed to housing price are identified and some justifications based on the previous studies will be given. The second part of this section investigates the performance of the models in predicting the housing price from the testing sample set. All statistical hypotheses were conducted at a 0.05 level of significance.

4.1 Attributes significantly contributed to housing price

A total of seven M_pULFR models for Petaling District and its six sub-regions are constructed. The transacted housing price is correlated with eight housing attributes and a time factor. The time factor is used to differentiate the transaction time of repeat-sales houses. All housing attributes were normalized before constructing the models. Table III shows the estimated parameters for M_pULFR and MR models and their performance measures in Petaling Jaya, Shah Alam, Sungai Buloh, Puchong, Seri Kembangan and Subang Jaya.

Attributes	Petaling Jaya				Shah Alam			
	M_PULFR	p-val.	MR	p-val.	M_PULFR	p-val.	MR	p-val.
Constant	−3201.36	–	−417.13	5.45E–13	−13120.67	–	−230.78	1.6E–06
Lot size	9264.85	0.0000	1037.80	6.5E–241	24775.06	0.0000	4242.50	0.0000
Tenure type	−2094.88	0.0000	143.76	2.08E–03	−1487.71	0.0000	76.28	0.0533
Time to expiry	2621.80	0.0000	340.39	2.2E–08	1829.54	0.0000	235.85	3.02E–06
Terrace type	7739.19	0.0000	435.73	4.2E–38	10727.04	0.0000	−113.90	9.15E–08
Bedrooms no.	8216.49	0.0000	50.76	0.0417	4746.85	0.0000	−69.35	5.51E–06
Building size	6637.34	0.0000	1811.63	2.4E–272	3007.68	0.0000	1395.17	0.0000
Mall distance	−9766.77	0.0000	−285.56	1.49E–78	9972.66	0.0000	−174.10	2.84E–66
Supermarket Dis.	−14091.05	0.0000	−112.54	3.73E–13	7397.08	0.0000	63.79	2.12E–14
Transaction date	3365.03	0.0000	576.33	0.0000	2675.77	0.0000	373.72	0.0000
Model MSE	1.84E–07		40856.55		1.38E–05		18058.46	
R^2	0.9999997		0.7171		0.9999998		0.8184	
Attributes	Sungai Buloh				Puchong			
	M_PULFR	p-val.	MR	p-val.	M_PULFR	p-val.	MR	p-val.
Constant	−3249.39	–	98.47	0.4902	−4127.02	–	−285.74	4.11E–06
Lot size	3403.91	0.0000	462.04	3.1E–21	7457.99	0.0000	1075.49	0.0000
Tenure type	−755.68	0.0000	−64.23	0.5260	−4291.02	0.0000	128.48	9.24E–03
Time to expiry	968.11	0.0000	40.50	0.7417	4892.61	0.0000	282.37	9.57E–06
Terrace type	2292.34	0.0000	−46.05	0.0949	8773.64	0.0000	−80.70	7.18E–04
Bedrooms no.	3629.91	0.0000	0.64	0.9823	4274.03	0.0000	−20.94	0.09039
Building size	1306.05	0.0000	685.58	2.77E–46	4179.32	0.0000	1678.66	0.0000
Mall distance	−2916.83	0.0000	−85.04	0.0706	−18731.48	0.0000	−120.26	1.02E–80
Supermarket Dis.	5844.27	0.0000	−113.54	7.91E–03	12261.81	0.0000	−30.05	2.38E–04
Transaction date	4311.73	0.0000	305.90	1.24E–54	2177.06	0.0000	454.05	0.0000
Model MSE	6.82E–08		8819.52		8.06E–08		16206.46	
R^2	0.9999988		0.8483		0.9999993		0.7721	
Attributes	Seri Kembangan				Subang Jaya			
	M_PULFR	p-val.	MR	p-val.	M_PULFR	p-val.	MR	p-val.
Constant	23068.01	–	−373.75	2.39E–30	−18373.48	–	−391.00	0.0406
Lot size	10511.04	0.0000	884.09	2E–140	11579.74	0.0000	1886.40	0.0000
Tenure type	8575.89	0.0000	288.68	3.09E–23	−11245.92	0.0000	110.24	0.5309
Time to expiry	−25149.91	0.0000	349.01	9.13E–24	12352.46	0.0000	264.15	0.1664
Terrace type	2434.93	0.0000	−56.49	1.1E–08	18214.99	0.0000	338.86	4.57E–25
Bedrooms no.	3989.68	0.0000	−10.33	0.4253*	9869.07	0.0000	75.43	4.98E–05
Building size	1998.76	0.0000	661.59	0.0000	3610.81	0.0000	1004.98	0.0000
Mall distance	−54273.64	0.0000	−26.91	1.66E–04	−14776.55	0.0000	−168.48	3.02E–50
Supermarket Dis.	5545.67	0.0000	62.48	2.77E–25	13421.37	0.0000	94.46	1.52E–13
Transaction date	1020.14	0.0000	311.30	0.0000	1792.68	0.0000	563.98	0.0000
Model MSE	7.34E–04		6542.79		9.14E–05		14913.70	
R^2	0.999995		0.7260		0.9999996		0.8121	

Table III. Estimated parameters for MpULFR model and MR model in six selected sub-regions

In Table III, it is shown that the proposed M_PULFR model has a better fitting ability in the training sample set as compared to the MR model. For example, the MSE and R^2 generated from the M_PULFR model for Petaling Jaya are 1.84E-07 and close to 1.0, respectively. This is much better than the MR model where its MSE is 40856.55 and R^2 is 0.7171. A similar result is observed for other sub-regions and the Petaling district.

4.1.1 Petaling jaya. All housing attributes are significant determinants of the housing prices in Petaling Jaya with p-values smaller than 0.05. Both M_pULFR and MR models show that lot sizes and main building sizes have a positive impact on housing prices. Buyers are willing to pay more for a larger lot and main building sizes which is also indicated in the studies from Pashardes and Savva (2009) and Owusu-ansah (2012). In the study of Ooi *et al.* (2014), freehold housings are preferable compared to leasehold housings. This finding is further supported by the M_PULFR model, but the MR model shows a positive relationship between housing prices and leasehold housings. The contradiction may due to a very high negative linear correlation ($\rho = -0.9906$) between tenure type (x_2) and time to expiry (x_3) that contributed to the existence of multicollinearity in MR model with variation inflation factors (VIF) of 79.21 and 78.31 for the variables x_2 and x_3 respectively which have seriously affected the estimation of MR model.

M_PULFR and MR models show that house buyers prefer a house with a longer length of the residential lease, and they are willing to pay more to own a house with more bedrooms. Distance to the nearest amenities such as shopping mall and supermarket have a negative impact on the housing prices in Petaling Jaya. This can be interpreted as the house buyers in Petaling Jaya are more willing to invest in the houses that have better accessibility and convenience. A similar remark was also noted in Kam *et al.* (2016). However, as Rosiers *et al.* (1996) has pointed out, the impact of the distance to nearest amenities on housing prices is ambiguity where these attributes have contributed either repulsion or attraction effect.

4.1.2 Shah alam. All housing attributes are significant determinants of the housing prices in Shah Alam except the tenure type (p-value = 0.0533) by the MR model. M_pULFR model produces positive relationships between housing prices and all housing attributes, except tenure type.

In particular, M_pULFR model shows that the distance to the nearest shopping mall and supermarket have a significant positive relationship with the housing prices in Shah Alam. This indicates that house buyers in Shah Alam are less likely to spend more money to buy housings that are near to shopping malls and supermarkets. This phenomenon can be explained because the earlier development of Shah Alam prior to 2000 is solely based on a unique identity of a Malay City with no entertainment outlets (City Declaration by Shah Alam City Council as cited in Wikipedia, 2017a, 2017b). Besides, the repulsion effect of shopping malls and supermarkets on housing prices in Shah Alam may also due to a not preferable noise and air pollutions, and traffic congestions (Tse and Love, 2000).

In contrast, the MR model shows the distance to the nearest shopping mall has a negative impact, while the distance to the supermarket has a positive impact on the housing prices. The most perplexity outputs from the MR model are the terrace type (number of floors) and the number of bedrooms, where both show negative relationships with housing prices. This is contradicted to Babawale and Adewunmi (2011), Owusu-ansah (2012), and Pashardes and Savva (2009) where they concluded that the number of bedrooms contributes a positive impact to housing prices.

4.1.3 Sungai Buloh, Puchong and Subang Jaya. The M_pULFR model produces the same results in Sungai Buloh, Puchong and Subang Jaya with all housing attributes are significant. Most of the attributes have a positive impact on the housing prices except tenure type and distance to the nearest shopping mall. Unlike Petaling Jaya and Shah Alam, house buyers in Sungai Buloh, Puchong and Subang Jaya preferred to invest in houses that are nearer to a shopping mall, but further to the supermarket. It is believed that supermarket usually causes serious traffic congestion and insufficient parking spaces to the neighborhood, while the shopping mall provides in-house parking lots.

For the MR model, five out of nine housing attributes and the intercept are statistically not significantly contributed to the housing prices in Sungai Buloh. Four significant housing attributes are lot size, building size, distance to the nearest supermarket and the transaction date. The MR model indicates that distance to the nearest supermarket has a negative impact on housing prices. This is contradicting with M_pULFR findings.

In the Puchong area, the MR model indicates only the number of bedrooms is not a significant determinant for housing prices. Among the significant determinants, it is observed that lot size, tenure type, time to expiry, building size and transaction date have a positive relationship with housing prices; while terrace type, distance to nearest shopping mall and distance to nearest supermarket have a negative impact.

There are two insignificant determinants in Subang Jaya using the MR model, namely tenure type (p-value = 0.5309) and time to expiry (p-value = 0.1664). From all the significant determinants, only distance to the nearest shopping mall is negatively related to housing prices.

4.1.4 Seri Kembangan. Consistently, the M_pULFR model resulted in a significant impact of all housing attributes for Seri Kembangan. The main different obtained from Seri Kembangan as compared to other regions is that the tenure type has a positive impact, while the time to expiry has a negative impact on the housing prices in Seri Kembangan. This means that the leasehold houses have higher market value than the freehold houses, while the shorter the lease term, the higher the housing prices in Seri Kembangan. This is contradicted to the previous studies in different areas where freehold housings are preferable as compared to leasehold housings (Ooi *et al.*, 2014), and newer houses with longer lease term are preferable as compared to older houses (Chiang *et al.*, 2015; Clapp and Giaccotto, 1998; Tse and Love, 2000).

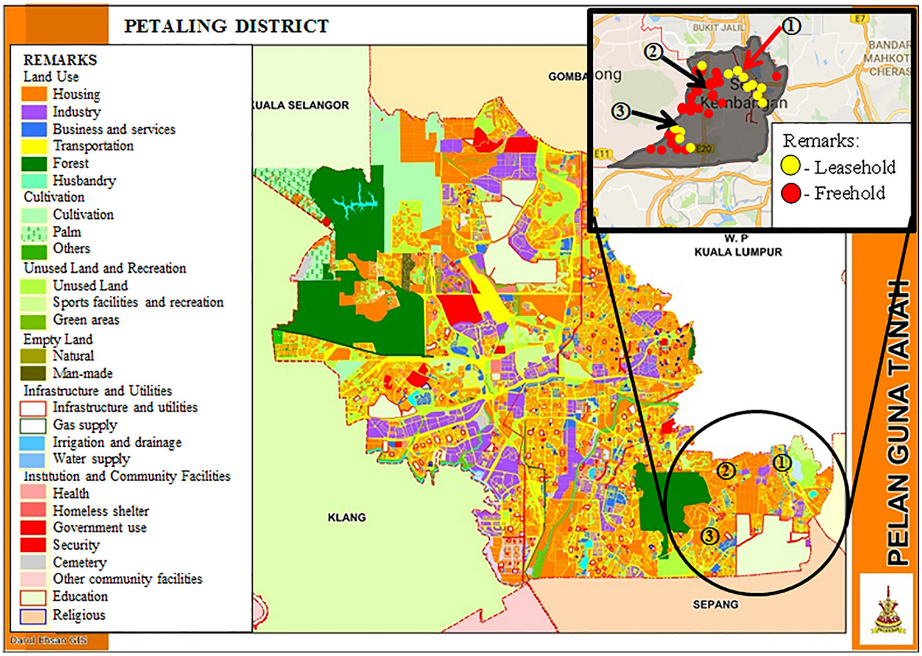
These phenomena are further investigated and could be explained when considering the geographical locations of the terrace houses in Seri Kembangan as shown in Figure 2. It is observed that leasehold housings are mainly located at areas that nearby industrial area or business centre (labeled as ① and ②) as compared to freehold housings (labeled as ③). Industrial and business areas are usually seen as attractive force to the housings market due to more job opportunities in these areas, and hence it causes the prices of the leasehold housings are higher as compared to freehold housings.

Similar to Puchong, the MR model shows a negative relationship between the terrace type and housing prices in Seri Kembangan. The only possible explanation is that residents in Seri Kembangan preferred single-storey houses as compared to double or higher number of storey houses. The only insignificant housing attribute using the MR model is the number of bedrooms.

4.1.5 Petaling district. The previous sub-sections identified significant housing attributes for each region within Petaling District. It is also interested to compare if the housing attributes changed at the district level. Table IV indicates the estimated parameters for M_pULFR and MR model for the whole Petaling District.

When considering a larger residential area, the M_pULFR model remains outperformed than the MR model with much smaller MSE and higher R^2. M_pULRF is also more consistent where it produces the same set of significant housing attributes for sub-regions and district level. In general, house buyers in Petaling District are more likely to spend more in exchange for housings with a larger lot and building sizes, freehold tenure, newer houses, more floors, more bedrooms and good accessibility to shopping mall and supermarket.

On the other hand, the terrace type is not significant under the MR model while it reveals that the housing prices increased if there are a fewer number of bedrooms. Same to Shah Alam, this observation is contradicted to the previous studies (Babawale and Adewunmi, 2011; Owusu-ansah, 2012; Pashardes and Savva, 2009).



Source: Google Maps

Figure 2. Land use of Petaling District prepared by Unit DEGIS Pejabat and Tanah Petaling (2014) and geographical position of housings in Seri Kembangan

	All-region			
Attributes	M_PULFR	*p*-val.	MR	*p*-val.
Constant	10660.79	–	−363.46	1.85E − 36
Lot size	32433.17	0.0000	4107.73	0.0000
Tenure type	−2510.63	0.0000	167.15	3.67E − 14
Time to expiry	3232.74	0.0000	374.33	2.86E − 36
Terrace type	10711.51	0.0000	−16.66	0.2036
Bedrooms no.	9968.89	0.0000	−119.98	1.18E − 29
Building size	5077.58	0.0000	1897.75	0.0000
Mall distance	−11291.13	0.0000	−249.59	0.0000
Supermarket Dis.	−74566.31	0.0000	−93.11	1.13E − 60
Transaction date	2752.30	0.0000	460.77	0.0000
Model MSE	1.94E − 05		27102.99	
R^2	0.9999997		0.7268	

Table IV. Estimated parameters for MpULFR model and MR model in Petaling district

4.2 Predictions of housing prices

In the previous section, we have seen that the M_pULFR model outperformed the MR model in estimating parameters with higher consistency and accuracy for the training sample. In this section, we will compare the predictive accuracy of the M_pULFR and MR models when applied to the testing sample. Table V summarizes the performance of housing prices prediction for Petaling District and its sub-regions.

Table V.
Prediction accuracy of housing prices using MpULFR model and MR model

	Petaling Dsitrict		Shah Alam		Puchong		Petaling Jaya	
	M_PULFR	MR	M_PULFR	MR	M_PULFR	MR	M_PULFR	MR
With <5% difference	22.85%	14.67%	26.13%	17.21%	27.12%	14.67%	20.71%	16.29%
With <10% difference	42.12%	28.16%	45.70%	33.67%	47.97%	29.59%	40.16%	31.08%
With <30% difference	83.25%	67.98%	85.62%	74.66%	88.01%	72.66%	82.07%	70.37%
MSE for difference <30%	6962.23	9810.61	5603.87	7866.59	4796.68	6552.06	11916.71	15483.80
MSE (testing samples)	19034.64	27191.46	13281.47	21956.83	10753.67	18665.33	28421.70	38256.35
Nearest *h* (best)	5		5		4		4	
Smallest *h*	1		1		1		2	

(*continued*)

	Subang Jaya		Seri Kembangan		Sungai Buloh	
	M_PULFR	MR	M_PULFR	MR	M_PULFR	MR
With <5% difference	25.72%	23.59%	25.81%	18.38%	24.72%	12.36%
With <10% difference	50.36%	44.11%	47.17%	35.73%	40.45%	24.72%
With <30% difference	90.53%	87.01%	87.10%	78.64%	79.78%	66.29%
MSE for difference <30%	6155.91	6631.29	2383.94	3102.83	3575.97	4266.40
MSE (testing samples)	14016.46	14305.99	6027.63	6847.15	18509.61	20621.78
Nearest *h* (best)	5		5		7	
Smallest *h*	3		3		3	

Table V.

In term of prediction ability, the M_PULFR model remains outperformed MR model in all cases. There is a significant reduction in MSE values for those cases with less than 30 per cent difference from actual housing prices (Figure 3). There are 20.71 to 27.12 per cent of the predicted housing prices using the M_PULFR model are very close to the actual housing prices with only less than 5 per cent of differences. The differences between predicted and actual housing prices with less than 5 per cent using the MR model range from 12.36 to 23.59 per cent.

Similar results obtained when we consider the predicted and actual housing prices are of 10 and 30 per cent differences. In Sungai Buloh, for example the M_pULFR model have 79.78 per cent cases with less than 30 per cent difference from actual value, while the MR model only achieved 66.29 per cent. The highest percentage achieved is in Subang Jaya where 90.53 and 87.01 per cent cases with less than 30 per cent difference from actual housing prices from the M_pULFR model and MR model, respectively.

The overall performance of the models in the testing sample is given by the MSE. The MSE values for the M_pULFR model are much smaller than the MR model for all regions. In Petaling District, the MSE are 19034.64 and 27191.46 for M_pULFR and MR models respectively. The overall MSE values in the pair of (M_pULFR, MR) for the respective sub-regions are (13281.47, 21956.83) for Shah Alam, (10753.67, 18665.33) for Puchong, (28421.70, 38256.35) for Petaling Jaya, (14016.46, 14305.99) for Subang Jaya, (6027.63, 6847.15) for Seri Kembangan and (18509.61, 20621.78) for Sungai Buloh.

There is a significant reduction in MSE values for those cases with less than 30 per cent difference from actual housing prices. In Petaling District, 83.25 per cent of the cases have less than 30 per cent of the difference between predicted and actual housing prices and its MSE is 6962.23, which is much lower than the MSE (19034.64) for the entire testing sample. This means that a sizable portion of MSE, that is 12072.41 in Petaling District is caused by the remaining 16.75 per cent of the cases. The same situations happened to the sub-regions for M_pULFR and MR models.

The h nearest houses showed in Table V are the numbers of houses with the most similar housing attributes needed to achieve the best predictions by resulting the smallest MSE for the M_PULFR model. In other words, there exists a nearest housing price for every h-similar-

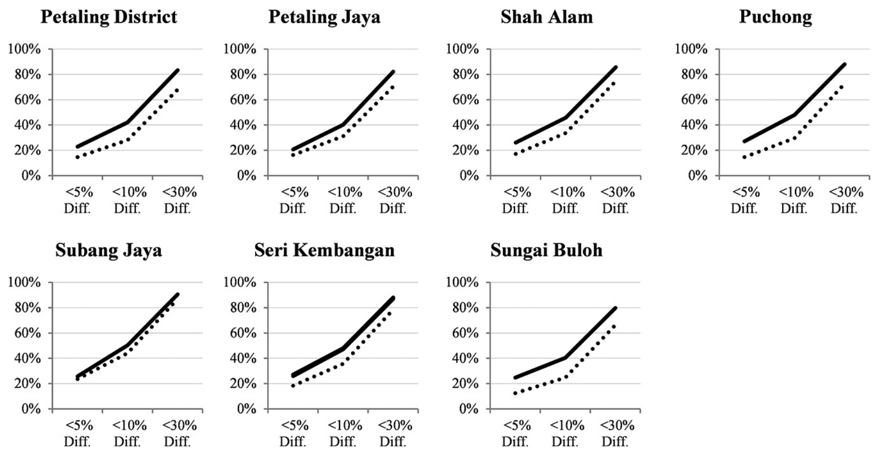


Figure 3.
Percentage comparisons within 5, 10 and 30% difference between actual housing price and predicted housing price using MpULFR Model (dotted line) and MR Model (smooth line)

housing. As Sungai Buloh has the largest value of h = 7, this implies that the housing market in Sungai Buloh is relatively more volatile as it needs seven similar housings to achieve best predictions while the other regions need only four or five similar housings.

Besides, the smallest h in Table V indicates the numbers of similar housings required by the M_PULFR model to achieve a better prediction as compared to the MR model. In general, the M_PULFR model needs at least three nearest houses to achieve results that will outperform MR model.

5. Discussion and concluding remarks

This study has applied a new M_PULFR model to investigate the influence of a set of housing attributes on the terrace-house prices in Petaling district, Malaysia. The proposed model able to improve the prediction ability of HPM and provides a good indicator of the volatility of the housing market after considering the error aspects of the attributes. The housing attributes that's not significant under other prediction model such as MR may actually strong determinants of housing price, reducing errors associated with the estimation approach. This study had successfully discovered the impact of various housing attributes to housing price at Petaling Jaya's different sub-regions and its district level. The findings confirmed that house buyers in Petaling district have similar preferences on the common housing attributes. It was also found that there were localized factors that affect the housing prices such as Islamic concept city that impede entertainment development in Shah Alam and more strategic location of the leasehold houses in Seri Kembangan. In addition, our analysis indicated that the housing market in Sungai Buloh is relatively more volatile than other sub-regions. Despite the encouraging results from the proposed model, it remain a challenge to generalize the implications nationwide as the study only uses a large actual housing transaction records that represent a segment of Malaysian urban population. Future studies might need to aggregate other localized factors before applying the model (in this case, Islamic concept city in Shah Alam) at other areas. Based on these observations, it is suggested that policymakers, property developers and investors could consider the impact of the housing attributes and the volatility of the housing prices in their future development planning, especially on urban area. On the other hand, the potential house buyers could use the model to evaluate if a house is overpriced by considering the proposed housing attributes.

References

Adyan, N.A., Hilman, T., Ruth, E.F. and Wayan, F.M. (2017), "Modeling house price prediction using regression analysis and particle swarm optimization", *International Journal of Advanced Computer Science and Applications*, Vol. 8 No. 10, pp. 323-326.

Babawale, G.K. and Adewunmi, Y. (2011), "The impact of neighbourhood churches on house prices", *Journal of Sustainable Development*, Vol. 4 No. 1, pp. 246-253.

Case, K.E. and Shiller, R.J. (1987), "Prices of single-family homes since 1970: new indexes for four cities", *New England Economic Review*, September/October, pp. 45-56.

Case, K.E. and Shiller, R.J. (1989), "The efficiency of the market for single-family homes", *The American Economic Review*, Vol. 79 No. 1, pp. 125-137.

Central Bank of Malaysia (2013), "Financial stability and payment systems report 2012: developments in the housing market and implications on financial stability".

Chang, Y.F., Mohd Rijal, O. and Abu Bakar, S.A.R. (2010), "Multidimensional unreplicated linear functional relationship model with single slope and its coefficient of determination", *WSEAS Transactions on Mathematics*, Vol. 9 No. 5, pp. 295-313.

Chau, K.W. and Chin, T.L. (2002), "A critical review of literature on the hedonic price model and its application to the housing market in Penang", *Housing Science and Its Application*, Vol. 27 No. 2, pp. 145-165.

Chiang, Y.H., Peng, T.C. and Chang, C.O. (2015), "The nonlinear effect of convenience stores on residential property prices: a case study of Taipei, Taiwan", *Habitat International*, Vol. 46, pp. 82-90, doi: 10.1016/j.habitatint.2014.10.017.

Chin, T.L. and Chau, K.W. (2003), "A critical review of literature on the hedonic price model", *International Journal for Housing and Its Applications*, Vol. 27 No. 2, pp. 145-165.

Choong, W.C., Pan, W.Y. and Chang, Y.F. (2018), "Multiple un-replicated linear functional relationship model and its application in real estate", *AIP Conference Proceedings 1974, 040016*, doi: 10.1063/1.5041690.

Clapp, J.M. and Giaccotto, C. (1995), "Which real estate price index is right for morgage-backed securities?", *Real Estate Finance*, Vol. 12 No. 2, pp. 41-47.

Clapp, J.M. and Giaccotto, C. (1998), "Residential hedonic models: a rational expectations approach to age effects", *Journal of Urban Economics*, Vol. 44 No. 3, pp. 415-437, doi: 10.1006/juec.1997.2076.

Clark, D.E. and Herrin, W.E. (2000), "The impact of public school attributes on home sale price in California", *Growth and Change*, Vol. 31 No. 3, pp. 385-407.

Clauretie, T.M. and Neill, H.R. (2000), "Year-round school schedules and residential property values", *The Journal of Real Estate Finance and Economics*, Vol. 20 No. 3, pp. 311-322.

Francis, A. (2004), *Business Mathematics and Statistics*, 6th ed., Cengage Learning EMEA, Canada.

Gabriel, S. (1984), "A note on housing market segmentation in an Israeli development town", *Urban Studies*, Vol. 21 No. 2, pp. 189-194.

Hamid, A. and bin Mar, I. (2006), *Basic Aspects of Property Market Analysis*, Universiti Teknologi Malaysia, Skudai, ISBN 983-52-0399-7.

Herath, S. and Maier, G. (2010), "A framework for analysing house prices using time, space and quality criteria", *Proceedings of the 17th Annual European Real Estate Society Conference*, pp. 1-22.

Ismail, S. and Macgregor, B.D. (2005), "Hedonic modelling of housing markets using geographical information system (GIS) and spatial statistics: a case study of glasgow", Scotland. PhD thesis, University of Aberdeen, Scotland.

Jiang, L., Phillips, R.C.B. and and Yu, J. (2014), "A new hedonic regression for real estate prices applied to the Singapore residential market", Cowles Foundation Discussion Paper, No. 1969, Yale University:

Kam, K.J., Chuah, S.Y., Lim, T.S. and Ang, F.L. (2016), "Modelling of property market: the structural and locational attributes towards Malaysian properties", *Pacific Rim Property Research Journal*, Vol. 22 No. 3, pp. 203-216, doi: 10.1080/14445921.2016.123436.

Ketkar, K. (1992), "Hazardous waste sites and property values in the state of New Jersey", *Applied Economics*, Vol. 24 No. 6, pp. 647-659.

Lee, L.L. (2014), "The inflation-hedging characteristic of malaysian residential property", *International Journal of Housing Markets and Analysis*, Vol. 7 No. 1, pp. 61-75.

Lee, C.L. and Ting, K.H. (2011), "Linkage between malaysian houisng prices, property companies and stocks", *Pacific Rim Property Research Journal*, Vol. 17 No. 2, pp. 287-312, doi: 10.1080/14445921.2011.11104329.

Lian, H.H. and Smyth, R. (2012), "Regional house prices and the ripple effect in Malaysia", *Urban Studies*, Vol. 50 No. 5, pp. 895-922, doi: 10.1177/0042098012459582.

Lim, P.I. and Chang, Y.F. (2018), "Preference of residential typologies of urban Malaysian, the malaysian institute of planners", *Planning Malaysia Journal*, Vol. 16 No. 3, pp. 171-181.

Megbolugbe, I.F. (1989), "A hedonic index model: the housing market of jos, Nigeria", *Urban Studies*, Vol. 26 No. 5, pp. 486-494.

Monson, M. (2009), "Valuation using hedonic pricing models", *Cornell Real Estate Review*, Vol. 7, Article 10, pp. 62-73.

NAPIC (2015), "The malaysian house price index: explanatory notes", pp. 51-59, available at: http://napic.jpph.gov.my/portal/web/guest/main-page

Neelam, S. and Kiran, G. (2018), "Valuation of house prices using predictive techniques", *International Journal of Advances in Electronics and Computer Science*, Vol. 5 No. 6, pp. 34-40.

Norshazwani, A.R., Mohd, L. and Abdul, J.O. (2013), "The revisited of Malaysian house price index", *Melaka: Proceedings International Conference of Technology Management, Business and Entrepreneurship 2012*, pp. 656-667, available at: http://eprints.uthm.edu.my/3381/

Ong, T.S. (2013), "Factors affecting the price of housing in Malaysia", *Journal of Emerging Issues in Economics, Finance and Banking*, Vol. 1 No. 5, pp. 414-429.

Ooi, J.T.L., Le, T.T.T. and Lee, N.J. (2014), "The impact of construction quality on house prices", *Journal of Housing Economics*, Vol. 26, pp. 126-138, doi: 10.1016/j.jhe.2014.10.001.

Ottensmann, J.R., Payton, S. and Man, J. (2008), "Urban location and housing prices within a hedonic model", *Journal of Regional Analysis and Policy*, Vol. 38 No. 1, pp. 19-35.

Owusu-ansah, A. (2012), "Examination of the determinants of housing values in urban Ghana and implications for policy makers", *Ghana: African Real Estate Society Conference*, pp. 1-16.

Pashardes, P. and Savva, C.S. (2009), "Factors affecting house prices in Cyprus: 1988-2008", *Cyprus Economic Policy Review*, Vol. 3, pp. 3-25.

Rosiers, F.D., Lagana, A., Thériault, M. and Beaudoin, M. (1996), "Shopping centres and house values: an empirical investigation", *Journal of Property Valuation and Investment*, Vol. 14, pp. 41-62, doi: 10.1108/14635789610153461.

Shiau, H.K., Ismail, N.W. and Chin, L. (2018), "The sources of house price changes in Malaysia", *International Journal of Housing Markets and Analysis*, Vol. 11 No. 2, pp. 335-355, doi: 10.1108/IJHMA-04-2017-0039.

So, H.M., Tse, R.Y.C. and Ganesan, S. (1996), "Estimating the influence of transport on house prices: Evidence from Hong Kong", *Journal of Property Valuation and Investment*, Vol. 15 No. 1, pp. 40-47.

Tan, Y.K. (1999), "An hedonic model for house prices in Malaysia", *Proceedings of the PRRES Conference*, *Kuala Lumpur*, available at: www.prres.net, pp. 1-15.

Tse, R.Y.C. and Love, P.E.D. (2000), "Measuring residential property values in Hong Kong", *Property Management*, Vol. 18 No. 5, pp. 366-374, doi: 10.1108/02637470010360669.

Tyrvainen, L. (1997), "The amenity value of the urban forest: an application of the hedonic pricing method", *Landscape and Urban Planning*, Vol. 37 Nos 3/4, pp. 211-222.

Unit DEGIS Pejabat, Daerah and Tanah, Petaling (2014), Pelan gunatanah semasa Daerah Petaling [Current land use of Petaling District].

Wikipedia (2017a), "Petaling district", available at: https://en.wikipedia.org/wiki/Petaling_District

Wikipedia (2017b), "Shah alam", available at: https://en.wikipedia.org/wiki/Shah_Alam

Wong, W.L., Chin, L. and Seow, S.K. (2019), "Housing prices in peninsular Malaysia: supported by income, foreign inflow or speculation?", *International Journal of Housing Markets and Analysis*, Vol. 12 No. 3, pp. 424-441, doi: 10.1108/IJHMA-01-2018-0001.

Yusof, A.M. and Ismail, S. (2012), "Multiple regressions in analysing house price variations", *Communications of the Ibima*, doi: 10.5171/2012.383101.

Further reading

Ismail, S. and Bryan, M. (2006), "Hedonic modelling of B197 housing markets using geographical information system and spatial statistics: Glasgow", *International Real Estate Research Symposium 2006*, *Scotland, Kuala Lumpur*.

Joint Center for Housing Studies (2016), "Chapter 6. The state of the nation's housing 2016: housing challenges", Harvard University, pp. 31-36, available at: www.jchs.harvard.edu/sites/jchs.harvard.edu/files/son_2016_200dpi_ch6.pdf

Kamarudin, N. Ismail, S. Ali, H.M. Sipan, I. and Mahamud, R. (2008), "Modelling of the property market: the malaysian experience", available at: http://eprints.utm.my/5591/

Matignon, R. (2007), *Chapter 2 Explore Nodes. Data Mining Using SAS Enterprise Miner*, Wiley-Interscience, p. 100.

Wong, R. (2017), "Turning Hong Kong's housing challenge into a housing solution", available at: www.scmp.com/business/article/2066745/turning-hong-kongs-housing-challenge-housing-solution

Ong, T.S. and Chang, Y.S. (2013), *Macroeconomic Determinants of Malaysian Housing Market*, Vol. 1 No. 2, pp. 119-127.

Corresponding author
Yun Fah Chang can be contacted at: changyf@utar.edu.my