A REVIEW ON THE PERFORMANCE OF HOUSE PRICE INDEX MODELS: HEDONIC PRICING MODEL VS ARTIFICIAL NEURAL NETWORK MODEL

Aida Julia Aminuddin
Nurul Hana Adi Maimun

1 Department of Real Estate, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, Johor, Malaysia, (E-mail: ajulia3@graduate.utm.my)
2 Centre for Real Estate Studies, Institute for Smart Infrastructure and Innovative Construction, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, Johor, Malaysia, (Email: nurulhana@utm.my)

Abstract: Most of the countries around the globe including Malaysia using the Hedonic Pricing Model (HPM) to build a national house price index. However, HPM seems to not reflect the current market trend due to its non-linearity and multicollinearity. Hence, one of the alternative techniques that have been applied is an artificial neural network to improve the reliability of forecasting information. The present article conducted a systematic literature review on the artificial neural network in forecasting local house price index. This study integrated multiple research designs and the review was based on the publication standard, namely ROSES (RepOrting Standard for Evidence Syntheses). This study selected articles through various databases such as Scopus, Emerald, Science Direct, SpringerLink, Web of Science, and Google Scholar. This review has five main themes namely 1) The Malaysian House Price Index (MHPI); 2) Hedonic Pricing Model; 3) Main problem of hedonic price function; 4) Artificial Neural Network (ANN), and 5) Performance of ANN. The findings explained the performance of artificial neural network in forecasting house price index based on previous study. This study benefits to the expansion of academic knowledge on property valuation in Malaysia where this research explores for improvement in the valuation sector using a reliable approach that is ANN model.

Keywords: House price index, forecasting, accuracy, hedonic pricing model, artificial neural network

Introduction

The current Hedonic Pricing Model (HPM) is widely used in most countries for predicting a national house price index. Despite adopting HPM in real estate valuation research (Bender et al., 2000), it seems to be a dominant method used by the experts occasionally in practice (McCluskey et. al, 2013). This statement supported by Selim (2009), who argued that HPM approaches not only used in theory but also has been widely used in practice. It is because the model shows its simplicity and straightforwardness (Chin et. al, 2003). However, this approach is said to have ineffectively functioned in terms of its predictive accuracy and reliability of the
developed model (Limsombunchai et. al, 2004; Lin and Mohan, 2011). It has been criticized due to exposure to the violation of traditional models which are nonlinearity, multicollinearity, and heteroscedasticity (Peterson and Flanagan, 2009; Kilpatrick, 2011; Antipov and Pokryshevskaya, 2012). However, the fact that the variables in the hedonic model are non-linear contradicts the classical model assumptions, which are argued to impact price prediction accuracy. Decision-making by several parties, including policymakers and developers, would have a negative impact on the limitations of the existing hedonic model used.

Recently, there is increasing attention to adopt Artificial Intelligence (AI) to forecast house prices. According to Yalpir (2014), real estate researchers and practitioners are encouraged to adopt AI appraisal techniques to enhance the accuracy of valuation estimation. Artificial Intelligence (AI), through Artificial Neural Network (ANN), was identified to be able to address the shortcomings of HPM such as nonlinearity and multicollinearity (Tabales et.al, 2013). Despite the growing interest in the ANN model around the globe, there are very limited studies on the application of ANN price modelling in the housing sector in this country. However, most of its exploration and adoption in this sector have been emanated only from developed countries (Do and Grudnitski, 1992; Tay and Ho, 1992; Lam et.al, 2009; Selim, 2009; McCluskey et.al, 2013; amongst others).

According to Limsombunchai et.al (2004), there are various published results of the adoption of the ANN technique in the real estate research domain. Mooya (2015) and Abidoye and Chan (2016) added that less attention was given to the adoption of ANN technique in the real estate domain, especially in developing countries. The current Malaysian ANN modelling literature showed some limitations in which the employment of aggregated data rather than disaggregated data in house price modelling besets the usefulness of the model in predicting the actual property value.

Meanwhile, the weakness of commonly used approaches such as HPM has led to a shift towards the ANN technique. Past studies have proven that ANN technique is a better substitution to the commonly used - HPM approach, in property appraisal (Taffese, 2006). A conceivable reason for this is that the ANN technique produces estimations that is highly precise, can handle the non-linear relationships between attributes factor (inner layer) and house prices (output layer), has a self-learning ability which allows it to analyze an almost incomprehensible amount of data (Mohd Radzi et.al, 2012).

This technique also able to handle outliers, objective and user-friendly (Borst, 1991; Tay and Ho, 1992; Cechin et.al, 2000; Mora Esperanza, 2004). In addition, studies by Abidoye (2016) have reported that by adopting the ANN approach, their datasets for testing purposes can be processed at a greater speed and the results produced are more reliable and accurate. This study integrated multiple research designs and the review was based on the publication standard, namely ROSES (RepOrting Standard for Evidence Syntheses). This study collecting articles through various databases such as Scopus, Emerald, Science Direct, SpringerLink, Web of Science and Google Scholar.
Literature Review
There are five main themes that will be discussed in this section. This section will give overview on these themes regarding the models used for house price indices and its performance based on previous study in some countries.

The Malaysian House Price Index (MHPI)
The MHPI first came into existence in the year 1997 and was formerly published by the National Institute of Valuation (INSPEN) and subsequently by the National Property Information Centre (NAPIC). To date, a total of 118 series of quarterly and annual house price indices have been published covering a total of 41 districts/regions and 14 states in Malaysia (National Property Information Centre, 2009). Variables affecting house prices used in the construction of a house price index covers from locational, structural, and neighbourhood and are expressed in these types of attributes: floor level, land area, age of the building, distance from the nearest town centre, floor area/built-up area, type of house, interest/type of tenure, building quality, and neighbourhood classification.

The MHPI only produces a house price index at the district level where the information could be too general for some decision-makers. Although an aggregate index is easy to understand and compare with other economic indices, it is less meaningful due to “lumping” all housing units into one basket (National Institute of Valuation, 1996) and subject to bias (Watkins, 1999). Therefore, for house price indices to be accurate and meaningful, they should be estimated at the highest disaggregation level.

A good hedonic study disaggregates information at the neighbourhood level (Tse, 2002). Much research proved that the price prediction using disaggregated level is better than an aggregated level (Goodman, 1978; Forrest, 1991; Fletcher et al., 2000; Berry et al., 2002; and Goodman and Thibodeau, 2003). The variations in the market composition happen due to the differences of price exist dwelling with types and neighbourhoods (Forrest, 1991). Due to the difference in median prices between regions, failing to acknowledge the existence of housing submarkets would cause the regression models to be biased (Watkins, 1999) and unstable (Maguire et al, 2016).

While constructing house price indices, disaggregation of data along geographic lines is important (Goh et al, 2012). By creating a house price index, a series of sub-indices based on a certain neighbourhood or location would be extremely valuable. As a result, the most disaggregated dataset should be used to estimate price indices. Previous studies based in Australia such as Costello et al (2011) and Hatzvi and Otto (2008) observed house price variations across metropolitan areas. Meanwhile, a study based in Johor Bahru, Malaysia by Adi Maimun (2011) and Adi Maimun et al (2012) investigates the effects of disaggregation using a small set of data containing various neighbourhoods. These neighbourhoods are compared with the aggregated data provided by MHPI. The results show that there are obvious gaps between MHPI and the neighbourhood-level trend lines. This implies the need to disaggregate information at the highest possible level to avoid loss of information.

Hedonic Pricing Model (HPM)
Hedonic pricing modelling implies that a commodity (for example, houses) is made up of combinations of distinct components or qualities and that customers will buy a bundle of them to maximise their utility functions (Limsonbunchai, 2004). Consequently, Suriatini et al., (2009) recorded that the hedonic pricing model is used in economic goods price analyses. This
price theory has been used widely both in real estate appraisal and real estate economics due to its capability to control the characteristics of properties or property attributes (Selim, 2009; Adi Maimun, 2010). Whilst hedonic analysis can do house price forecasting using aggregate or disaggregate data, however, it is found that hedonic price coefficients of a few attributes are unstable. Adi Maimun (2010) stated that in adopting the hedonic model, sales price on various house characteristics needs to be regressed. The hedonic method is reported to be used in various countries (Ghorbani. S. and Afghae, 2017; Abidoye and Albert, 2018 and others), especially for house price valuation.

In the housing context, the hedonic method treats properties as composite products by regressing the prices on a vector of house features such as locational, structural and neighbourhood. The hedonic price of attributes, which is used for price estimate and prediction, is the marginal contribution of housing qualities arising from the hedonic model. This method addresses two critical concerns raised by Fleming and Nellis (1981) in constructing the house price index, namely, the heterogeneity of houses and the representativeness of the data in terms of house type, by considering properties as composite products. Therefore, it has been a very popular technique used in constructing house price indices (Bourassa et al., 2006; Pavlin, 2006; and Nappi-Choulet et al., 2007).

According to Osipova (2004), United States conducted many works on the hedonic price indices. However, many other countries also apply the hedonic method to construct national house price index such as United Kingdom (Hoesli and MacGregor, 2000), Finland (Koev, 2003), Germany (Linz and Behrmann, 2004), Hong Kong (Lum, 2004), Taiwan (Osipova, 2004), Switzerland (Bourassa et al., 2006), and Norway (Ljones, 2010). The typical hedonic price function is formulated as follows:

\[ y_i = p^* (z_i) + \varepsilon_i; \ i = 1, ..., n \]

where;
- \( y_i \) is the house price.
- \( p \) is the regression surface or coefficient.
- \( z_i \) is the locational, structural and neighbourhood variables.

**Main Problem in Hedonic Price Function**

There are a few problems associated with the hedonic price function namely nonlinearity, multicollinearity and heteroskedasticity (Tabales et al., 2013). Heteroskedasticity happens when the variance of the disturbance of the hedonic model is unequal (Fletcher et al (2000)). Heteroskedasticity, which usually took place in cross-sectional models may cause biased hedonic estimates, leading to unreliable hypothesis testing (Studenmund, 2006).

Multicollinearity is present when the model demonstrates a strong relationship between independent variables by having relationship value equal to or greater than 0.70, low t-values, Variance Inflation Factor of 5 or 10 and above with a tolerance less than 0.20 or 0.10 (O’Brien, 2007), Condition Number above 30, or non-orthogonal variables in the Farrar-Glauber Test. Multicollinearity causes high variances and specification errors (Kennedy, 2003).

This model has been criticized due to exposure to the violation of traditional models which are non-linearity, multicollinearity, and heteroscedasticity (Peterson and Flanagan, 2009; Kilpatrick, 2011; Antipov and Pokryshevskaya, 2012). Furthermore, it is worried that the use
of the HPM method as property price modelling, particularly for the construction of the Malaysian House Price Index (MHPI) and mass appraisal, would lead to inaccurate house price forecasting. As price-influencing variables frequently differ in coefficients, the pre-specified transformed variables do not capture the volatility of the housing market. In all related house variables such as locational, structural and neighbourhood, these variations are obvious. However, the fact that the variables in the hedonic model are non-linear, contradicts the classical model assumptions, which are argued to impact price prediction accuracy. Decision-making by several parties, including policymaker and developers, would have a negative impact on the limitations of the existing hedonic model.

The movements of house prices in Malaysia are unpredictable. Thus, our housing industry has been urged to enhance the current model, hence, the drawbacks of the current model can be improved. The improvement of HPM must be done so that it will also enhance the accuracy in forecasting house prices. Many research shows by implementing ANN as an advanced approach resulting in a more accurate forecast of house prices. The use of advanced approaches to predicting house price fluctuations is unavoidable at this age and period. (Waziri, 2010).

Additionally, Limsombunchai et al. (2004) and Lin & Mohan (2011) stated that HPM is unable to effectively capture the non-linear relationship that exists between house prices and property attributes despite inaccuracy issues. These weaknesses give a negative impact on house price prediction. Therefore, an artificial neural network model has been offered as the alternative for forecasting house prices.

**Artificial Neural Network (ANN)**

ANN has been applied in various application areas including classification or pattern recognition, forecasting and modelling. According to Sharda (1994), one of the successful applications was in the forecasting area. The first significant reason is that ANN has a higher learning efficiency. According to McCluskey et.al (2013), the greater improvement in the learning efficiency of ANN is due to efforts from many researchers. For instance, Lee et al. (2001) used ESP (error saturation prevention) method to speed up the learning efficiency and escape from local minimum; Ampazis et al. (2001) help feed-forward networks to minimize the training time by developed a constrained optimization algorithm. While two practices were introduced by Yin et al. (2003) which are to alleviate the extrapolation limitation of ANNs and to overcome the over-training limitation of ANNs. Their results have greatly improved the accuracy for ANN modelling and forecasting non-linear time series. This statement was supported by McCluskey (1996) and Ge et al. (2002) which the errors of the coefficient adjustment can be minimized by using ANN model for price estimation. Due to the lack of a restrictive assumption of linearity about the nature of data relationships and functional form, the ANN technique has been employed as the focus of several real estate studies for its capacity to assess relationships in complicated non-linear data sets (Nguyen & Cripps, 2001; Zurada et. al, 2006; Lin & Mohan, 2011). Their findings showed that ANN could accomplish tasks in the same way as conventional, parametric statistical techniques like regression and classification, but with greater prediction performance.

Thus, according to Peterson and Flanagan (2009), regression-based techniques are one of the most popular in the real estate sector, although they have major non-linearity issues. This claim is based on their research, which found that when compared to linear hedonic pricing models, this technique shows high accuracy in prediction performance. Moreover, the ANN generates
significantly lower pricing errors, have greater pricing precision out-of-sample, and extrapolate better from more volatile pricing environments. In this regard, ANN as a low-cost system is perceived to provide more robust outcomes in terms of model misspecification and the deformity in the measurement of explanatory variables (Peterson & Flanagan, 2009).

The applications of these numerous non-parametric and ANN techniques are competing for widespread use and uptake within the appraisal sector in terms of valuation model accuracy. The goal of ANN development was to mimic human cognitive performance in theory. A neural network is, at its most basic level, a processing device implemented as an algorithm that employs numerous simple processing units in the form of a network (Spellman, 1999). As a result, it is just a system that adjusts the weights of the nodes between processing elements based on patterns in a dataset. According to Warner and Misra (1996), this weighting adjustment leads to increased prediction and generalisation patterns, implying that the system is learning.

As a result, ANNs are digitised models that can not only handle non-linear functions but also learn from previously unknown relationships. They are adaptive, can cope with noisy and ambiguous data, and are incredibly computationally efficient once trained (Spellman, 1999). There is one proposed framework of ANN modelling for house prices forecasting. Based on this proposed framework, few criteria will be selected and transfer to phase 2 which the data will be processed to establish ANN dataset. In phase 3, the data will be distributed to another three chambers which known as training set, testing set, and validation set to undergo the next stage for input layer, hidden layer and finally output layer. At phase 4, the evaluation of the model takes part, and the results will show either the ANN will give a positive impact or not enhance the accuracy of forecasting house prices.

In relation to Government Policy, this study supports one of the aspirations of the National Transformation 2050 (TN50) as documented in the TN50- Aspirasi Malaysia, Rancangan Malaysia Kesebelas (RMK 11), National Key Result Areas, National Priority Areas, and National Policy on Industry 4.0 by utilising Artificial Intelligence in modelling house price indices. This is essential to prepare for future technological advancements and to remain competitive in the fast-changing world especially in property valuation.

**Artificial Neural Network (ANN)**

Table 1 shows the summary of the previous studies on the ANN in forecasting a house price index. This proved that the integration of artificial intelligence through ANN can enhance the accuracy of a house price index. Despite the growing interest in ANN model around the globe, there are very limited studies on the application of the ANN price modelling in the housing sector in this country. However, most of its exploration and adoption in this sector have been emanated only from developed countries (Do and Grudnitski, 1992; Tay and Ho, 1992; Lam et.al, 2009; Selim, 2009; McCluskey et.al, 2013; amongst others). To address the research gap, this paper aims to establish such a new model using ANN in forecasting house prices.
<table>
<thead>
<tr>
<th>Author/ Year</th>
<th>Study Area</th>
<th>Pricing Model</th>
<th>Variables</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nor, M. I., Masron, T. A., &amp; Gedi, S. Y. (2019)</td>
<td>Mogadishu, Somalia</td>
<td>Hedonic Pricing Model (HPM)</td>
<td>Artificial Neural Network (ANN)</td>
<td>Property attributes</td>
</tr>
<tr>
<td>Wang, L., Chan, F. F., Wang, Y., &amp; Chang, Q. (2017)</td>
<td>Singapore</td>
<td>Artificial Neural Network (ANN)</td>
<td>Property attributes</td>
<td>ANN is a useful tool in housing prices prediction and other financial applications. The price index movement can therefore be estimated, with a relatively small error.</td>
</tr>
<tr>
<td>Authors</td>
<td>Location</td>
<td>Model/Methods</td>
<td>Property attributes</td>
<td>Results/Findings</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Mohd Radzi, M. S.,</td>
<td>Malaysia</td>
<td>Multiple Regression Analysis (MRA), Artificial Neural Network (ANN)</td>
<td>Unemployment rate,</td>
<td>Neural network was a good alternative method in lieu of the traditional multiple regression analysis (MRA) for predicting house price. $R^2=99.32%$</td>
</tr>
<tr>
<td>Muthuverappan, C.,</td>
<td></td>
<td></td>
<td>population size and</td>
<td></td>
</tr>
<tr>
<td>Kamarudin, N., &amp;</td>
<td></td>
<td></td>
<td>household income</td>
<td></td>
</tr>
<tr>
<td>Mohammad, I. S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amri, S., &amp; Tularam, G.</td>
<td>Bathurst (NSW),</td>
<td>Multivariate Regression (MR), Neural Network (NN), Adaptive Neuro-Fuzzy (ANFIS)</td>
<td>Property attributes</td>
<td>Based on adjusted R squared, the Neural Networks prediction is showing better results than the fuzzy and MR. Adjusted R-square: 0.4536</td>
</tr>
<tr>
<td>A. (2012)</td>
<td>Australia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zurada, J., Levitan, A.</td>
<td>Louisville</td>
<td>Multiple Regression Analysis (MRA), Support Vector Machines Using Sequential</td>
<td>Property attributes</td>
<td>Combining neural networks and fuzzy logic produced results comparable to MRA. Employing a hybrid system might be a viable option. $R^2=90.90%$</td>
</tr>
<tr>
<td>S., &amp; Guan, J.</td>
<td>Kentucky</td>
<td>Minimal Optimization Regression (SVM-SMO), Additive Regression, M5P Trees,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2011)</td>
<td></td>
<td>Neural Networks (NNs), Radial Basis Function Neural Network (RBFNN), Memory-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Based Reasoning (MBR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selim (2009)</td>
<td>Turkey</td>
<td>Hedonic Pricing Model (HPM), Artificial Neural Network (ANN)</td>
<td>Property attributes</td>
<td>ANN is a better alternative model for house prices prediction in Turkey.</td>
</tr>
<tr>
<td>Pagourtzi et al.</td>
<td>Attica urban area</td>
<td>Multiple Linear Regression (MLR), Artificial Neural Network (ANN)</td>
<td>Property attributes</td>
<td>ANN is an efficacy model and able to deal with nonlinear relationships. ANN performs better than MLR due to nonlinearity es issue of the latter forecasting model.</td>
</tr>
<tr>
<td>(2007)</td>
<td>in Greece</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hedonic Price Model (HPM)</th>
<th>Artificial Neural Network (ANN)</th>
<th>Property attributes</th>
</tr>
</thead>
</table>

ANN model can overcome some of the problems related to the data patterns and the underlining assumptions of HPM. $R^2 = 90.00\%$

Based on the table above, the higher the coefficient of determination ($R^2$), the more accurate it is. Overall, the findings concluded that ANN is capable of forecasting highly accurate house prices as measured through $R^2$. By comparing the prediction performance between some models such as hedonic regression with ANN models, this study demonstrates that ANN can be a better alternative for prediction of the house prices.

**Conclusion**

An evaluation on selected empirical studies shows that the integration of ANN can overcome the shortcomings of current hedonic pricing model in forecasting house price index. The performance of ANN had been discussed and proved by research in this study area. This paper contributes to assists Valuation of Property and Services Department, Local Authorities, National Valuation Institute, Department of Property Valuation and Research in Private Organisations and other real estate stakeholders to apply the modelling framework in forecasting local house prices that are more reliable and accurate.

**Acknowledgement**

The authors would like to express their appreciation for the financial support awarded by the Ministry of Higher Education Malaysia through Fundamental Research Grant Scheme (FRGS/1/2019/SS08/UTM/02/7).

**References**


