Applied Soft Computing 65 (2018) 260–271

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Building real estate valuation models with comparative approach through case-based reasoning

I-Cheng Yeh [a,*], Tzu-Kuang Hsu [b]

[a] Department of Civil Engineering, Tamkang University, New Taipei City, 30067, Taiwan, ROC
[b] Department of International Business, Chung Hua University, Hsinchu, 30093, Taiwan, ROC

ARTICLE INFO

Article history:
Received 24 December 2016
Received in revised form
29 December 2017
Accepted 20 January 2018
Available online 31 January 2018

Keywords:
Quantitative
Comparative approach
Case-based reasoning
Real estate
Appraisal

ABSTRACT

The purpose of this study is to propose an innovative real estate valuation approach called Quantitative Comparative Approach, which can estimate correction coefficients to overcome the shortcomings of subjective decisions of correction coefficients of traditional comparative approach. The principle is to assume that the price per unit area of real estate is the average price per unit area of the particular circle of housing supply and demand multiplied by the product of several dimensionless adjustment coefficients of factors. The single regression models of these adjustment coefficients can be built with the stepwise decomposition regression analysis. Then the adjustment coefficients of comparative cases and target case can be estimated with these single regression models, and finally the correction coefficients can be estimated by dividing the adjustment coefficients of target case by those of comparative case. The empirical samples are collected from four circles of supply and demand, and are divided into four data sets. The empirical results show that the Quantitative Comparative Approach is more accurate than the two classical Hedonic price approaches, multivariate regression analysis and neural networks.

1. Introduction

Four modern approaches of real estate appraisals are income approach, cost approach, comparative approach, and hedonic price approach [1,2]. Although these existing approaches are all applicable, each of these is based on certain assumptions and has shortcomings along with its advantages.

The income approach looks at the sum of the real estate future cash (rent) flows as the price of the real estate. The shortcomings of income approach include high appraisal cost, time-consuming process, and that its estimated price is frequently deviate from the market price.

The cost approach assesses the price of the real estate by using the cost of land and house to build a new equivalent minus its depreciation of house. The main shortcoming of cost approach is that its estimated price does not consider marketability and profitability.

The comparative approach assumes that similar real estate should have similar benefit and price. It is therefore, also called the "relative appraisal method". The key point of this approach is to estimate the price of real estate by using the data of a set of similar or comparable cases traded on the market. In this approach, the prices of the comparative cases must be adjusted by the correction coefficients of factors before they are employed as the estimated price of target case. The correction coefficients often lack reasonable and objective quantitative methods to estimate their values, but rely on the appraisers' subjective evaluation to decide them. However, its estimated price is frequently close to the market price, it is generally accepted by appraisers. Nonetheless, the differences between appraisers and their subjectivity often make appraisal results inconsistent and bring up questions.

The hedonic price approach is the quantitative method of the comparative approach. Because the price of real estate is estimated by the regression model, it does not have the human bias which often happens in the conventional comparative approach. Furthermore, the regression model can decrease the appraisal cost, but it cannot handle the factors difficult to quantize [1,2]. So the hedonic price approach is often used by the industry to deal with the mass appraisal of real estate [2–6]. It is also used by the academies to play the role of benchmark to evaluate other approaches [7–11]. Compared with comparative approach, its main disadvantage is that it cannot provide real comparative cases for appraisers to valuate real estate price. In practice, comparative cases from real market trad-

* Corresponding author.
E-mail addresses: 140910@mail.tku.edu.tw (I.-C. Yeh), hsutk@chu.edu.tw (T.-K. Hsu).

https://doi.org/10.1016/j.asoc.2018.01.029
1568-4946/

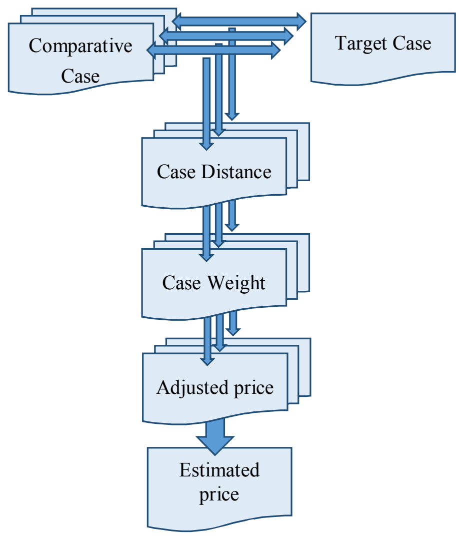

Fig. 1. The comparison approach.

ing are very important to persuade traders to accept the appraisal price.

The comparative approach estimates the price of target case through comparison, analysis, adjustment, and integration of the comparative cases. Therefore, the principle of comparative approach is the same as the Case Based Reasoning (CBR), which has emerged as an alternative to rule-based reasoning techniques for the design of knowledge-based systems [12]. The comparative approach focuses on the comparative adjustment of the market price created by the market supply and demand. The price from the comparative approach is called the comparative price. This price is the weighted average of the adjusted prices derived from the comparative cases which is similar to the condition of target case.

The comparative approach comprises four steps (see Fig. 1):

Step 1 Comparison. Calculate the distance between the comparative cases and target cases. The higher the similarity, the smaller the distance.

Step 2 Weighting. Calculate the weight of the comparative cases. The smaller the distance, the higher the weight of the comparative case.

Step 3 Adjustment. Adjust the price of the comparative cases with the correction coefficients of factors

$$Y_j = y_j \times \left(\prod_{i=1}^{m} K_i \right) \tag{1}$$

where Y_j is the adjusted price per unit area of comparative case j; y_j is the market price per unit area of comparative case j; K_i is the correction coefficient of the factor i.

Step 4 Integration. Integrate the weight of the comparative cases in Step 2 and the adjusted price of the comparative cases in Step 3 with the weighted average approach.

There are four procedures in the CBR methodology, that is, retrieval, revision, reuse, and retaining. In this study, the Step 1 (Comparison) and Step 2 (Weighting) are corresponding to the retrieval procedure. The Step 3 (Adjustment) can be regarded as the revision procedure. And the Step 4 (Integration) may play the role of the reuse procedure. After the real estate trading, it can be stored into the case base to implement the retaining procedure.

In the above steps, the correction coefficient of factors in step 3 is one of the keys to make the comparative approach be able to get the accurate estimated price for the target case. Traditionally, the correction coefficients of factors are usually subjectively determined by appraisers, but different appraisers got different correction coefficients and then obtained different estimated prices and brought up questions.

The purpose of this study is to propose an innovative approach called Quantitative Comparative Approach which can estimate correction coefficients to overcome the shortcomings of subjective decisions of traditional comparative approach.

2. Literature review

The hedonic price approach uses not only the regression analysis, but also artificial neural networks or other artificial intelligence approaches to establish the estimated price models of real estate [2–4,7–16]. For example, McCluskey, et al. [1] employed a sample of 2694 residential properties to compare geostatistical approach, artificial neural network (ANN) model, and the traditional linear hedonic pricing model for mass appraisal valuation accuracy. The findings demonstrate that ANNs can be shown to perform very well in terms of predictive power, and therefore valuation accuracy, outperforming the traditional multiple regression analysis and approaching the performance of spatially weighted regression approach. However, ANNs retain a 'black box' architecture that limits their usefulness to practitioners in the field.

Lin and Mohan [5] conducted a study to compare the prediction accuracy of the three most used models: multiple regression model, additive nonparametric regression, and ANNs. The three models were developed using the housing database of a town with 33,342 residential houses. The research confirmed that any of the three models can be used, with similar accuracy, for lower and medium–priced houses.

Zurada, et al. [6] conducted a comparative study in the real estate appraisal domain. Traditional multiple regression analysis, Support Vector Machines, and ANNs were compared under various simulation scenarios. The results indicate that ANNs perform well under most simulation scenarios.

Case Based Reasoning is a process of arriving at the solution of a new problem on the basis of the solutions of previously-solved similar problems. Gonzalez and Laureano-Ortiz [12] pointed out that the real estate appraisal is a domain characterized by having a single variable in its solution, that is the price of the real estate being appraised. This makes it different from most of other domains to which CBR has been applied. The solutions of the other domains may be some types of plans required the optimization of multiple objectives and satisfaction of multiple constraints. Moreover, the achievement of consistency is also essential in the real estate appraisal domain in which different appraisers may reach different answers even having the same data at their nature.

Musa, et al. [13] presented a Neural-CBR system for real estate valuation, which is based on a hybrid architecture that combines CBR techniques and Artificial Neural Networks. Their experimental results found that the system has higher satisfactory level of performance when compared with individual Artificial Neural Network and CBR systems. However, it is rather difficult to generate accurate fuzzy predicates. However, the 'black box' nature of ANN of the hybrid system limited its usefulness to appraisers in the field.

Bonissone and Cheetham [17] presented a CBR techniques with fuzzy predicates expressing preferences in determining similarities between subject and comparable properties. The fuzzy techniques are also used to estimate a confidence value qualifying such prediction. However, it is rather difficult to generate accurate fuzzy predicates.

In summary, although some hedonic price approaches such as ANNs may be more accurate than other approaches; however, compared with comparative approach, the ANNs approach cannot provide real comparative cases for appraisers to valuate real estate price. In real applications, comparative cases from real market trading are very important to persuade traders to accept the appraisal price. CBR solves new problems by retrieving cases of similar previously-solved problems and adapting their solutions to fit new situations. The case adaptation step is often the key to obtain solutions with high accuracy. Nevertheless, comparative approach lacks consistency and objectivity due to the appraisers' subjective evaluation on correction coefficients of appraisal factors. Therefore, in this paper we proposed a new method to objectively and automatically estimate the correcting coefficients.

3. Methodology

3.1. Principle

Firstly, we assume that the price per unit area of the target real estate is the average price per unit area of the particular circle of housing supply and demand multiplied by the product of several dimensionless adjustment coefficients of factors. Then the adjusted price of a comparative case is equal to

$$y_j = \bar{Y} \times k_{1j} \times k_{2j} \times k_{3j} \times k_{4j} \times \ldots \times k_{mj} \tag{2}$$

where $\bar{Y}$is the average price per unit area of the particular circle of housing supply and demand; k_{1j}, k_{2j}, k_{3j}, k_{4j},..., k_{mj} are the adjustment coefficients of the m factors for the comparative case j.

Similarly, the adjusted price for a target case is as follows.

$$y = \bar{Y} \times k_1 \times k_2 \times k_3 \times k_4 \times \ldots \times k_m \tag{3}$$

where $k_1, k_2, k_3, k_4, \ldots k_m$ are the adjustment coefficients of the m factors for the target case.

From Eq. (2) we get the average price per unit area of the particular circle of housing supply and demand in equation.

$$\bar{Y} = y_j \times \frac{1}{k_{1j}} \times \frac{1}{k_{2j}} \times \frac{1}{k_{3j}} \times \frac{1}{k_{4j}} \times \ldots \times \frac{1}{k_{mj}} \tag{4}$$

Substitute Eq. (4) into Eq. (3), we get the estimated price for the target case in Eq. (5).

$$\begin{aligned} y &= \left(y_j \times \frac{1}{k_{1j}} \times \frac{1}{k_{2j}} \times \frac{1}{k_{3j}} \times \frac{1}{k_{4j}} \times \ldots \times \frac{1}{k_{mj}} \right) \times k_1 \times k_2 \times k_3 \times k_4 \times \ldots \times k_m \\ &= y_j \times \frac{k_1}{k_{1j}} \times \frac{k_2}{k_{2j}} \times \frac{k_3}{k_{3j}} \times \frac{k_4}{k_{4j}} \times \ldots \times \frac{k_m}{k_{mj}} \end{aligned} \tag{5}$$

Let the correction coefficient of the factor i, K_i, equal the adjustment coefficient of the factor i for target case divided by the adjustment coefficient of the factor i for comparative case j in Eq. (6).

$$K_i = \frac{k_i}{k_{ij}} \tag{6}$$

where the adjustment coefficient of the factor i for the target case, k_i, and the adjustment coefficient of the factor i for the comparative case j, k_{ij}, are the function of the factor i. Then we can establish these functions (regression formulas) with the stepwise decomposition regression analysis.

Eq. (5) can be rewritten to Eq. (7)

$$y = y_j \times K_1 \times K_2 \times K_3 \times K_4 \times \ldots \times K_m \tag{7}$$

We introduce how to set up the regression model of these adjustment coefficients with the stepwise decomposition analysis and the whole steps for quantitative comparative approach in the next two subsections.

3.2. The quantitative model of the adjustment coefficients

We use the stepwise decomposition method to establish various adjustment coefficients as following steps.

(1) Sorting: use the sorting quantile method to measure the importance of factor.
(2) Decomposition: use the stepwise decomposition analysis to establish the regression model of adjustment coefficient for each factor according to the descending order of the importance of factor.

3.2.1. The sorting quantile method to measure the importance of factor

We use sorting quantile method to measure the importance of each factor to the price per unit area of real estate. Detailed steps are as follows:

(1) Sort the samples and divide them into ten portions according to the factor (independent variable).
(2) Calculate the average of dependent variable for each portion.
(3) Calculate the variance of dependent variable.

$$Var = \sum_{i=1}^{n} \left(Y_i - \bar{Y} \right)^2 \tag{8}$$

where Y_i is the average of dependent variable of the i-th portion; $\bar{Y}$is the average of dependent variable for all samples; n is the number of a portion, equal to 10.

The greater the variance derived from the factor, the greater the ability to explain the dependent variable of the factor, representing the greater the importance of the factor.

3.2.2. The stepwise decomposition regression analysis

The principle of the stepwise decomposition regression analysis is to assume that the price per unit area of real estate is the average price per unit area of the particular circle of housing supply and demand multiplied by the product of several dimensionless adjustment coefficients of factors.

$$Y_j = \prod_{i=1}^{m} k_{ij} \cdot \bar{Y} \tag{9}$$

where Y_j is the price per unit area of real estate for sample j; m is the number of factors; k_{ij} represents the adjustment coefficients of the factor i for the comparative case j; $\bar{Y}$is the average price per unit area of the particular circle of housing supply and demand.

The stepwise decomposition regression analysis starts from the most important factor, and decomposes and builds regression model of the adjustment coefficient for each factor, as shown in Fig. 2.

The importance of factor is estimated by the aforementioned **sorting quantile method**. The determinant factors of real estate include some attribute variables, including

- the distance to the nearest MRT station,
- the number of convenience stores in the living circle on foot,
- the house age,
- the transaction date, and
- the geographic coordinate consisting of the northing and the easting.

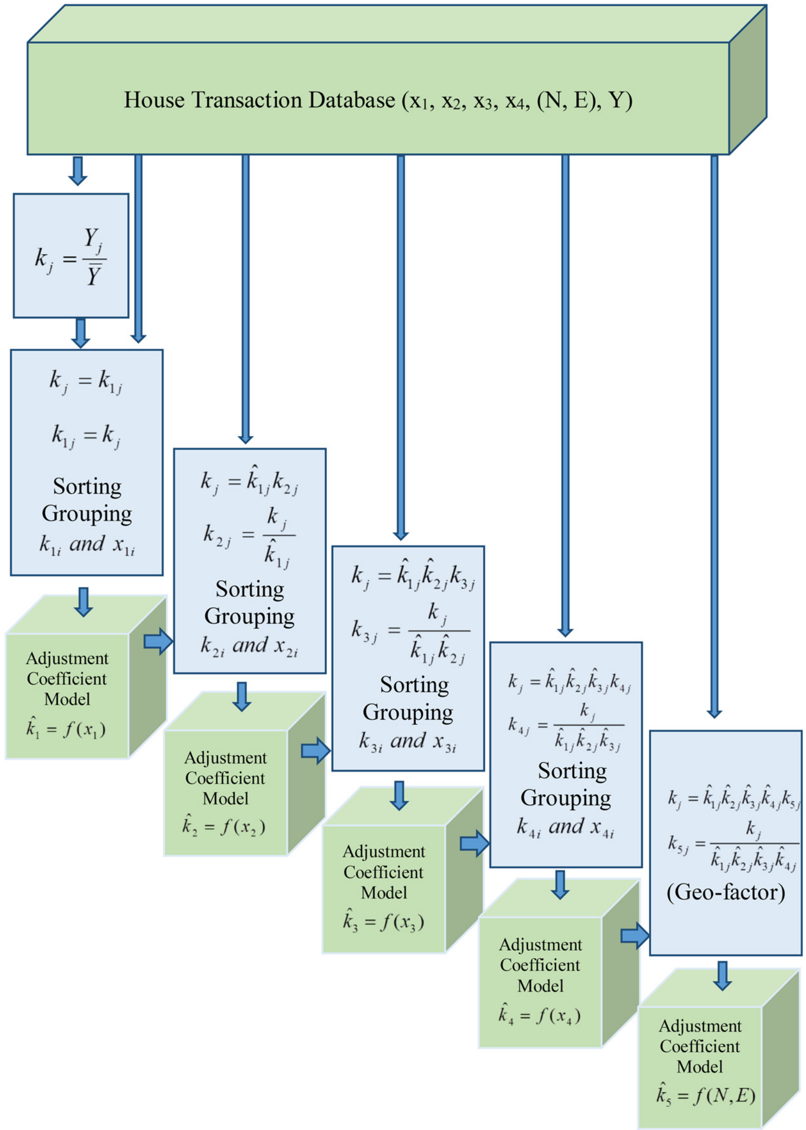

Fig. 2. The Stepwise Decomposion Regression Analysis.

The stepwise decomposition analysis to establish the regression model of adjustment coefficient for each factor according to the descending order of the importance of factor. Because the geographic coordinate consists of two variables, the northing and the easting, its importance cannot be estimated with the **sorting quantile method,** so in decomposition process, the factor of the geographic coordinate was put in the last.

The steps of the stepwise decomposition regression approach are as follow:

(1) Calculate the aggregate adjustment coefficient of each sample

$$k_j = \frac{Y_j}{\bar{Y}} \quad j = 1, 2, \ldots, N \tag{10}$$

where $\underline{k_j}$ is the aggregate adjustment coefficient of all factors for sample j; Y_j represents the price per unit area for sample j; and $\bar{Y}$is the average of the price per unit area for all sample of the supply and demand circle; and N is the number of sample in the supply and demand circle.

(2) Establish the model of the adjustment coefficient of the first important factor

- Calculate the adjustment coefficient of the factor for each sample

Suppose that the aggregate adjustment coefficient of each sample is totally determined by the adjustment coefficient of the most important factor. Then the aggregate adjustment coefficient equals the adjustment coefficient of the most important factor as the following Eq. (11).

$$k_j = k_{1j} \tag{11}$$

where k_{1j} = the adjustment coefficient of the most important factor, that is factor 1.

Therefore, Eq. (12) shows the adjustment coefficient of each sample.

$$k_{1j} = k_j \tag{12}$$

- Calculate the adjustment coefficient of the factor for each portion.

$$k_{1i} = \frac{\sum_{j \in i} k_{1j}}{numberofsampleintheportioni}$$

- Establish the regression model of the adjustment coefficient of the factor as Eq. (13).

$$\hat{k}_{1i} = f(a_1, b_1, x_{1i}) \tag{13}$$

where (a_1, b_1) are the parameters of the regression model.

We minimize the square errors of the difference of the forecast and actual adjustment coefficient to estimate (a_1, b_1) as Eq. (14).

$$MinE = \sum_{i=1}^{10} (\hat{k}_{1i} - k_{1i})^2 = \sum_{i=1}^{10} (f(a_1, b_1, x_{1i}) - k_{1i})^2 \tag{14}$$

(3) Establish the model of the adjustment coefficient of the second important factor

- Calculate the adjustment coefficient of the factor for each sample

Suppose that the aggregate adjustment coefficient of each sample is totally determined by the adjustment coefficients of the first and the second important factors in Eq. (15).

$$k_j = \hat{k}_{1j} k_{2j} \tag{15}$$

Then, in the both side we divided by $\hat{k}_{1j}$ and get Eq. (16).

$$k_{2j} = \frac{k_j}{\hat{k}_{1j}} \tag{16}$$

- Calculate the adjustment coefficient of the factor for each portion.

$$k_{2i} = \frac{\sum_{j \in i} k_{2j}}{numberofsampleintheportioni}$$

- Establish the regression model of the adjustment coefficient of the factor as Eq. (17).

$$\hat{k}_{2i} = f(a_2, b_2, x_{2i}) \tag{17}$$

where (a_2, b_2) are the parameters of the regression model.

The parameters can be estimated as Eq. (18).

$$MinE = \sum_{i=1}^{10} (\hat{k}_{2i} - k_{2i})^2 = \sum_{i=1}^{10} (f(a_2, b_2, x_{2i}) - k_{2i})^2 \tag{18}$$

(4) Establish the model of the adjustment coefficient of the rest important factor

Similarly, suppose that the aggregate adjustment coefficient of each sample is totally determined by the adjustment coefficients of the first three most important factors as Eq. (19).

$$k_{3j} = \frac{k_j}{\hat{k}_{1j} \hat{k}_{2j}} \tag{19}$$

And the regression model of the adjustment coefficient of the factor follows as Eq. (20).

$$\hat{k}_{3i} = f(a_3, b_3, x_{3i}) \tag{20}$$

Similarly, suppose that the aggregate adjustment coefficient of each sample is totally determined by the adjustment coefficients of the first four most important factors as Eq. (21).

$$k_{4j} = \frac{k_j}{\hat{k}_{1j} \hat{k}_{2j} \hat{k}_{3j}} \tag{21}$$

And the regression model of the adjustment coefficient of the factor follows as Eq. (22).

$$\hat{k}_{4i} = f(a_4, b_4, x_{4i}) \tag{22}$$

The regression model of the adjustment coefficient of all the rest factors can be built with the same procedure.

(5) Establish the model of the adjustment coefficient of the geographic coordinate

- Calculate the adjustment coefficient of the factor for each sample

Suppose that the aggregate adjustment coefficient of each sample is totally determined by the adjustment coefficient of the first four most important factors and the geographic coordinate as Eq. (23).

$$k_j = \hat{k}_{1j} \hat{k}_{2j} \hat{k}_{3j} \hat{k}_{4j} k_{5j} \tag{23}$$

Then the adjustment coefficient of the geographic coordinate follows as Eq. (22).

$$k_{5j} = \frac{k_j}{\hat{k}_{1j} \hat{k}_{2j} \hat{k}_{3j} \hat{k}_{4j}} \tag{24}$$

- Establish the regression model of the adjustment coefficient of the factor

The regression model of the adjustment coefficient of the geographic coordinate is a second-order polynomial model as Eq. (25).

$$\hat{k}_{5j} = c_0 + c_1 N_j + c_2 E_j + c_3 N_j^2 + c_4 E_j^2 + c_5 N_j E_j \tag{25}$$

where N_j, E_j are the northing and the easting values for the sample j; $(c_0, c_1, c_2, c_3, c_4, c_5)$ are the regression parameters.

We minimize the square errors of the difference of the forecast and actual adjustment coefficient to estimate regression parameters as Eq. (26).

$$Min \sum_{j=1}^{N} (\hat{k}_{5j} - k_{5j})^2 = \sum_{j=1}^{N} (f(c_0, c_1, c_2, c_3, c_4, c_5, N_j, E_j) - k_{5j})^2 \tag{26}$$

Where N = the number of sample in the data set.

3.2.3. The regression model

This study considers the following five regression models for the above the adjustment coefficient as Eqs. (27)–(31).

$$\text{Linear model } \hat{k}_i = a_i x_i + b_i \tag{27}$$

$$\text{Quadratic model } \hat{k}_i = a_i x_i^2 + b_i x_i + c_i \tag{28}$$

$$\text{Logarithmic model } \hat{k}_i = a_i \ln(x_i) + b_i \tag{29}$$

$$\text{Exponential model } \hat{k}_i = a_i \exp(b_i x_i) \tag{30}$$

$$\text{Exponential growth model } \hat{k}_i = a_i x_i^{b_i} \tag{31}$$

3.3. Quantitative comparative approach

3.3.1. The process of the quantitative comparative approach

The process of the quantitative comparative approach is as follows:

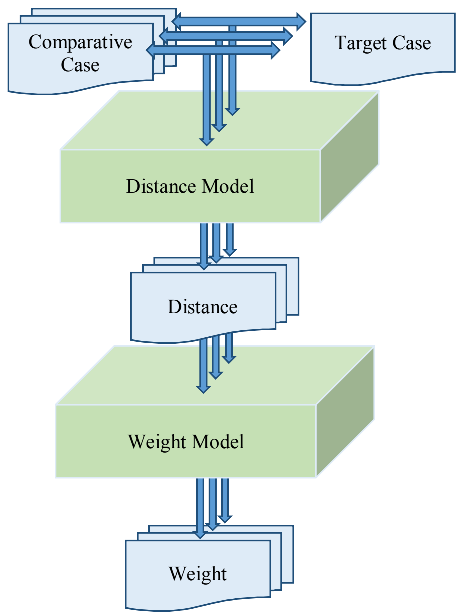

Fig. 3. The case distance model and case weight model.

(1) Calculate the case distance of the comparative case

The distance of the comparative case represents the difference between the target case and the comparative case j. The characteristics (factors) of the target case and those of the comparative case are substituted into the Eq. (32) to calculate its case distance.

$$D_j = \sqrt{\frac{\sum_{i=1}^{m} W_i \times \left(\frac{S_i - x_i^j}{std(x_i)}\right)^2}{\sum_{i=1}^{m} W_i}} \tag{32}$$

where m is the number of the characteristics variables; W_i is the weighted value of the characteristics variable i; S_i is the characteristic variable i of the target case; x_i^j = represents the characteristic variable i of the comparative case j; and $std(x_i)$ is the standard deviation of the characteristic variable i in the data set.

(2) Calculate the case weight of the comparative case

The case weight of the comparative case is used to estimate the price of the target case in the weight average method. The greater the distance between the target case and the comparative case j, the smaller the case weight of the comparative case. Therefore, the case weight of the comparative case can be measured by the following Eq. (33) (see Fig. 3).

$$F_j = \exp\left(-\left(\frac{D_j}{\delta}\right)^2\right) \tag{33}$$

where δ is the effect radius, δ > 0. As the case distance is close to 0, the case weight is close to 1; when the case distance is close to the effect radius, it is close to 0.368; and when the case distance is close to several times of the effect radius, it is close to 0.

When the effect radius is very small, the case weight of the comparative case which is the most similar to the target case is much greater than other comparative cases; hence, the result of the weighted average method may be controlled by the single comparative case most similar to the target case, like the nearest neighbor method. Instead, when the effect radius is very great, the case weights of all the comparative cases are all the same, close to 1; hence, the result of the weighted average method may be the average of the all comparative cases, like the simple average method. Both of the two extreme conditions cannot lead the weighted average method to attain the best results. Therefore, the effect radius should lead the case weights of the comparative case that are similar to the target case be appropriately greater than the ones of the comparative case that are not similar to the target case to lead the weighted average method to achieve the best results.

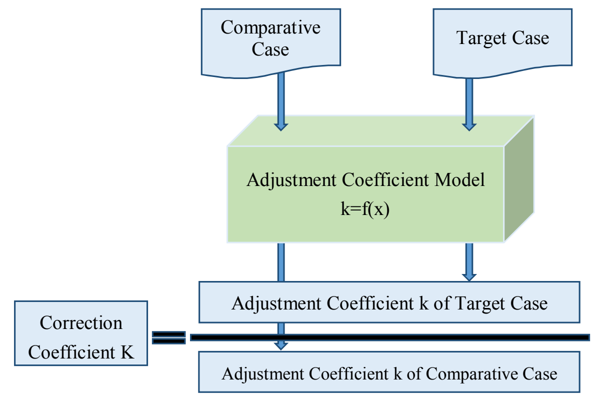

Fig. 4. The correction coefficient model.

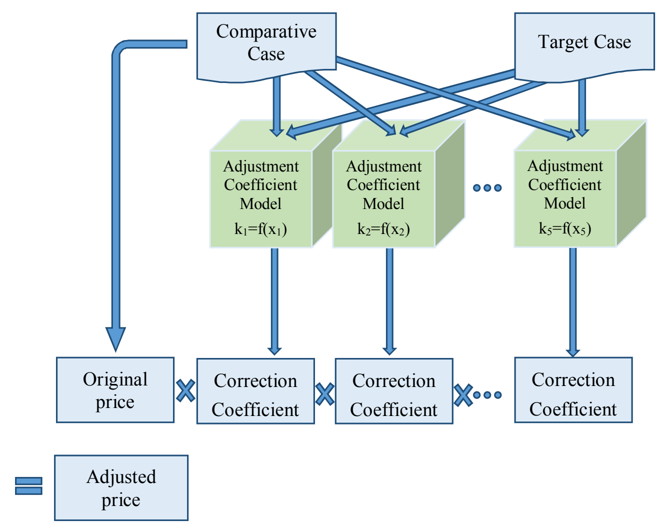

Fig. 5. The adjusted price model.

(3) The adjusted price of the comparative case

The adjusted price of the comparative case can be estimated by the following equation (see Figs. 4 and 5).

$$Y_j = y_j \times \left(\prod_{i=1}^{m} \frac{\hat{k}_i}{\hat{k}_{ij}}\right) \tag{34}$$

Where y_j is the original transaction price of the comparative case j; $\hat{k}_i$ is the adjustment coefficient of the factor i for the target case; and $\hat{k}_{ij}$ is the adjustment coefficient of the factor i for the comparative

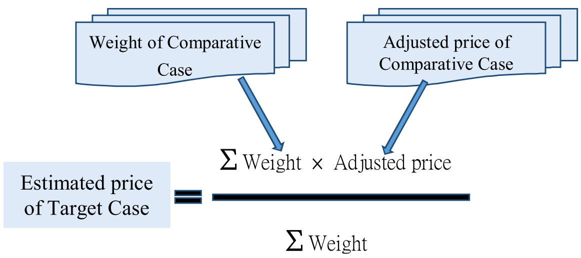

Fig. 6. The Weighted average model.

case j. These adjustment coefficients can be calculated with the regression models built with the stepwise decomposition method.

For example, the smaller the distance to the nearest MRT station, the higher the house price; hence, the greater adjustment coefficient. Suppose the distance to the nearest MRT station of the target case is greater than that of the comparative case; hence, the adjustment coefficient of the target case is smaller than that of the comparative case, for example, they are 0.9 and 1.1 respectively, then the correction coefficient of the factor is equal to 0.9/1.1 = 0.82.

(4) Calculate the integrated estimates

The price of the target case can be estimated by the weighted average method (see Fig. 6).

$$\hat{Y} = \frac{\sum_{j=1}^{N} F_j \times Y_j}{\sum_{j=1}^{N} F_j} \tag{35}$$

Where F_j is the case weight of the comparative case j; and N is the total number of the comparative case in the circle of housing supply and demand.

3.3.2. Parameter setting

In the above calculating, two types of parameters have to be set.

(1) The weights of the characteristics variables (appraisal factors)

When we calculate the case distance, the weights of the appraisal factors must be set. In this study we propose four methods.

Method 1: The weight of the factor is propotional to the root of variance of the dependent variable (house price) in the portions derived from sorting by the factor as Eq. (8).

Method 2: The weight of the factor is propotional to the variance of the dependent variable (house price) in the portions derived from sorting by the factor as Eq. (8)

Method 3: Similar to method 1, but use the average of the root of variance of the dependent variable (house price) in the portions derived from sorting by the factor of all the circles of housing supply and demand. Hence, all the circles of housing supply and demand use the same weights.

Method 4: Similar to method 2, but use the average of the variance of the dependent variable (house price) in the portions derived from sorting by the factor of all the circles of housing supply and demand.

The weight of the least important factor is set to 1, all the weights of the other factors are proportionally set. The weight of the geographic coordinate cannot be set as the above methods; therefore, it is set to a fixed value. The experiments showed that setting the weight of the geographic coordinate to 3 can obtain robust results.

(2) The effect radius

To calculate the case weight, the effect radius must be set. This study used the trial and error method to find the optimal radius, and the value 1.25, 1.5, 1.75, 2, 3, 5, and 100 were tried in this study.

4. Empirical design

4.1. Appraisal factors

The dependent variable is the residential housing price per unit area. Referring to the related researches [7–16], five appraisal factors (independent variables) were chosen as follows.

(1) The distance to the nearest MRT station

The distance to the nearest MRT station cannot be acquired directly from the public database. Instead, first the location coordinate (the northing and the easting) can be acquired by the address of the house through the google map, then the distance to each MRT station can be calculated by the coordinate of house and the coordinate of MRT station (in meter unit), then the distance to the nearest MRT station can be obtained by the minimization operation.

(2) The number of convenience stores in the living circle on foot

In Taiwan, the convenience stores are all over the urban and country and bring people living convenience. Therefore, the number of convenience stores in the living circle on foot may be an important factor of house price. We tried different distance to define the living circle on foot, and found setting the distance to 500 m can obtain robust results. This data cannot be acquired directly from the public database. Instead, first the location coordinate can be acquired by the address of the house with the google map, and the distance to the convenience stores can be calculated by the location coordinate of house and the convenience store, then the number of convenience stores in the living circle on foot (the distance < 500 m) can be counted.

(3) House age

The house age at the actual transaction (in year unit) not only affects depreciation of house but also the living quality of the indoor living circle.

(4) Transaction date

The house price is not only affected by the house but also the market. Therefore, the transaction date is an important factor of house price. It must be presented with a real number. For example, transaction date June 2012 is presented as 2012.5 and March 2013 is presented as 2013.25.

(5) Geographic coordinate

The location coordinate of house determines the distance to the downtown; hence, may determine the time and money cost to go shopping and to office.

4.2. Data collection

In this study the data set is collected from the public database of Ministry of the Interior during the period of June 2012 to May 2013 from two districts in Taipei City, and two districts in New Taipei City, totally four circles of supply and demand; hence, there

Table 1
Sample descriptive statistics of the four circles of supply and demand.

Independent variable	Circle of supply and demand	Min	Max	Mean	Standard deviation	Number
Distance to the nearest MRT station (meter)	Sindian	23.38	6488.02	1083.89	1262.11	414
	Danshuei	106.20	6636.80	1264.12	994.60	388
	Wunshan	37.87	2120.11	741.40	474.66	418
	Beitou	10.07	2509.31	589.08	507.37	743
Number of the convenience store in the living circle on foot	Sindian	0.00	10.00	4.09	2.95	414
	Danshuei	0.00	10.00	2.09	2.03	388
	Wunshan	0.00	8.00	3.53	1.93	418
	Beitou	0.00	10.00	3.62	2.65	743
House age (year)	Sindian	0.10	43.80	17.71	11.39	414
	Danshuei	0.10	34.90	8.16	7.17	388
	Wunshan	0.10	45.20	17.86	12.18	418
	Beitou	0.10	85.70	16.93	14.49	743
Transaction date (A.D.)	Sindian	2012.67	2013.58	2013.15	0.28	414
	Danshuei	2012.50	2013.42	2012.93	0.28	388
	Wunshan	2012.67	2013.58	2013.11	0.28	418
	Beitou	2011.33	2013.33	2012.73	0.38	743

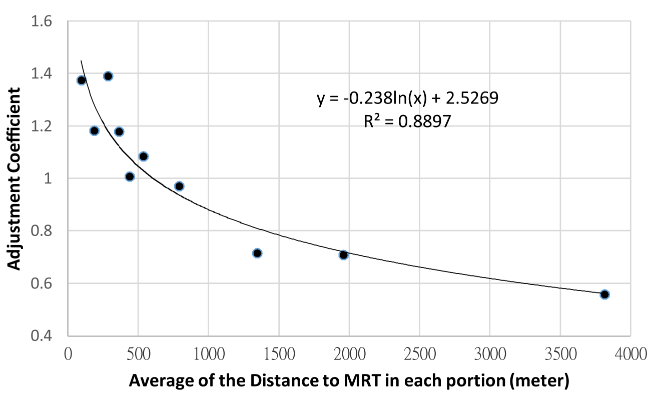

Fig. 7. The optimal regression model, logarithmic model, of the adjustment coefficient of the distance to the nearest MRT station with the Sindian data set.

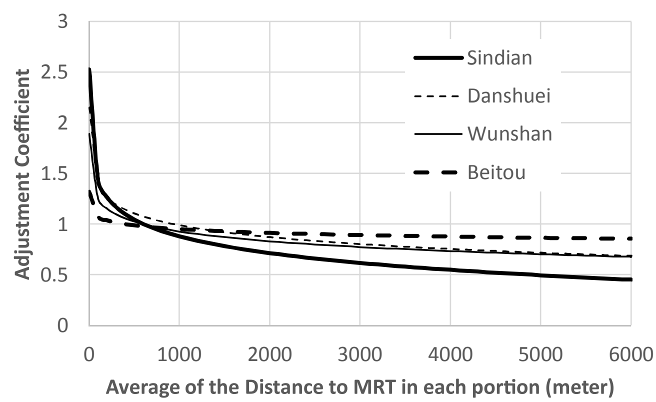

Fig. 8. The optimal regression model, logarithmic model, of the adjustment coefficient of the distance to the nearest MRT station with four data sets.

are four data sets in this study. Table 1 display the sample numbers and descriptive statistics of the independent variables. Each of the four data sets was randomly split into the training data set (2/3 samples) and the testing data set (1/3 samples). The training data sets were employed to build the quantitative models of adjustment coefficients. The testing data sets were employed to evaluate the accuracy of the results of the quantitative comparative approach.

5. Empirical results

5.1. *The quantitative model of adjustment coefficient*

Using the **sorting quantile method** and regression analysis, the factors ranking by their importance from high to low, as well as their optimal regression models of adjustment coefficient are listed as follows:

- Number of convenience stores in the living circle on foot (k2): linear model
- House age (k3): logarithmic model
- Transaction date (k4): linear model

For example, Fig. 7 shows the optimal regression model, logarithmic model, of the adjustment coefficient of the distance to the nearest MRT station with the Sindian data set, and Fig. 8 shows the adjustment coefficient regression models of the distance to the nearest MRT station with the four data sets. As the distance is longer than 500 m, the effect of the distance to the nearest MRT station becomes much smaller than that shorter than 500 m. As expected, because they are dimensionless models, then they may have some generality across various circles of supply and demand.

In this study, we used these above sequence and models to stepwise decompose and build adjustment coefficient regression models for the four circles of supply and demand. Table 2 show these regression models.

(1) Distance to the nearest MRT station

The factor can explain over 70% of the variance of the house price of three of the four circles of supply and demand. Overall, the distance of the nearest MRT station is the most important factor and its adjustment coefficient is inversely proportional to the distance of the nearest MRT station as expectation.

(2) Number of convenience stores in the living circle on foot

The factor can explain over 40% of the variance of the house price of three of the four circles of supply and demand. Overall, its adjustment coefficient is proportional to the number of convenience stores in the living circle on foot as expectation.

(3) House age

The factor can explain over 30% of the variance of the house price of three of the four circles of supply and demand. Overall, its adjustment coefficient is inversely proportional to the house age as expectation.

Table 2
Adjustment coefficient regression models.

circle of supply and demand	Coefficient a	Coefficient b	Coefficient of determination	Root of mean square error
Distance to the nearest MRT station				
Sindian	−0.23843	2.527609	0.8820	0.081076
Danshuei	−0.16818	2.153550	0.7625	0.059454
Wunshan	−0.13948	1.891902	0.7027	0.039520
Beitou	−0.05326	1.320662	0.3319	0.033849
Number of the convenience store in the living circle on foot				
Sindian	0.009452	0.955591	0.2830	0.018765
Danshuei	−0.038770	1.079642	0.6454	0.030437
Wunshan	0.033735	0.878961	0.4558	0.048586
Beitou	0.021259	0.922298	0.6376	0.017543
House age				
Sindian	−0.06408	1.159766	0.5238	0.036351
Danshuei	−0.02480	1.039141	0.1399	0.054020
Wunshan	−0.03891	1.096332	0.3719	0.034063
Beitou	−0.02584	1.042212	0.3931	0.056939
Transaction date				
Sindian	0.21170	−20.62480	0.7283	0.013932
Danshuei	0.07991	−7.14398	0.2700	0.027994
Wunshan	0.02667	−1.72339	0.0789	0.010950
Beitou	−0.05797	6.90521	0.1287	0.032050

Table 3
Weights of characteristic variables.

Method	circle of supply and demand	Distance to the nearest MRT station	Number of the convenience store in the living circle on foot	House age	Transaction date
Method 1	Sindian	3.7	2.9	2.3	1.0
	Danshuei	2.4	1.7	1.8	1.0
	Wunshan	3.9	3.5	3.7	1.0
	Beitou	1.3	1.4	1.8	1.0
Method 2	Sindian	13.4	8.2	5.5	1.0
	Danshuei	5.9	2.9	3.2	1.0
	Wunshan	15.4	12.0	13.6	1.0
	Beitou	1.6	1.9	3.3	1.0
Method 3		2.8	2.4	2.4	1.0
Method 4		9.1	6.2	6.4	1.0

(4) Transaction date

The explanation ability of the factor is unstable among the four circles of supply and demand. However, the factor can explain over 70% of the variance of the house price of the Sindian circle of supply and demand. To explore the reason, we checked the price index of these four circles, and found that during the studied period, Sindian district has a very big rising of price index through the period and has the greatest increase, follows by Danshuei district, Beitou district, and Wenshan district, which is consist with the coefficients of determination in Table 2.

5.2. The results of quantitative comparative approach

The regression models of the adjustment coefficients are the core of the quantitative comparative approach. The weights of the characteristic variables to calculate the case distance also have great impacts. The weights of characteristic variables determined by the above mentioned four methods are showed in the Table 3.

Using the quantitative comparative approach associated with these weights and different effect radii, the root mean squared errors (RMSE) of the unit area house price for the four data sets were obtained and shown in Table 4. For example, the actual and forecast house price per unit area of the Sindian district is shown in Fig. 9. The results are as follows:

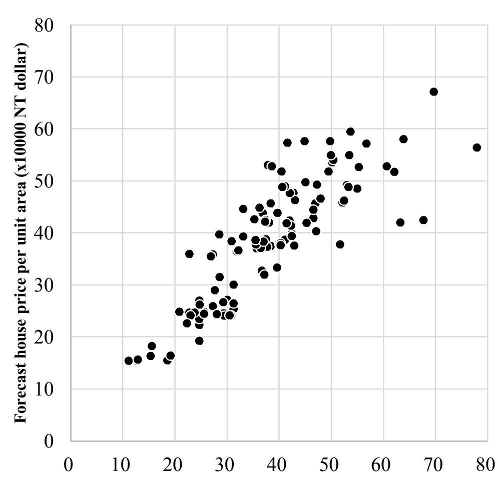

Fig. 9. The actual and predictive house price per unit area of the testing data set of Sindian.

(1) The weight of the appraisal factor

Comparing the minimum error of the four different methods determining weights of the appraisal factors, the errors for these four methods have only slight difference, but if we did not use the weights, the errors increased. Therefore, the weights of the

Table 4
The root mean squared error (RMSE) of the estimates per unit area house price.

circle of supply & demand	Radius								
	Weights	100	5	3	2	1.75	1.5	1.25	Min
Sindian	Method 1	7.98	7.95	7.77	7.68	7.78	7.80	7.82	7.68
	Method 2	7.98	7.83	7.80	8.23	8.30	8.35	8.45	7.80
	Method 3	7.98	7.95	7.81	7.73	7.74	7.82	7.87	7.73
	Method 4	7.98	7.92	7.91	8.12	8.27	8.35	8.39	7.91
Danshu	Method 1	4.48	4.43	4.28	4.04	3.93	3.81	3.91	3.81
	Method 2	4.48	4.26	3.98	3.70	3.65	3.63	3.66	3.63
	Method 3	4.48	4.37	4.16	3.88	3.78	3.70	3.77	3.70
	Method 4	4.48	3.98	3.66	3.66	3.65	3.63	3.65	3.63
Wunshan	Method 1	7.87	7.45	6.93	6.89	6.94	7.01	7.06	6.89
	Method 2	7.87	7.08	6.92	7.32	7.40	7.51	7.67	6.92
	Method 3	7.87	7.63	7.15	6.83	6.82	6.86	7.01	6.82
	Method 4	7.87	7.04	6.90	7.07	7.16	7.27	7.35	6.90
Beitou	Method 1	10.38	10.22	9.91	9.49	9.35	9.22	9.30	9.22
	Method 2	10.37	9.99	9.62	9.31	9.24	9.20	9.25	9.20
	Method 3	10.37	10.09	9.64	9.25	9.20	9.19	9.23	9.19
	Method 4	10.37	9.47	9.36	9.33	9.30	9.28	9.30	9.28

Table 5
The evaluation of the prediction with the quantitative comparative approach.

circle of supply & demand	20% error hit rate	10% error hit rate	R^2	RMSE	Avg. house price	RMSE/Avg. house price
Sindian	80.2	52.9	0.683	7.73	37.98	0.204
Danshuei	81.7	50.3	0.627	3.64	21.27	0.171
Wunshan	83.0	50.7	0.527	6.82	43.68	0.156
Beitou	67.0	38.4	0.492	9.19	41.20	0.223
Average	78.0	48.1	0.582	6.845	36.03	0.190

appraisal factors can improve the prediction of the quantitative comparative approach.

(2) The effect radius

The effect radii with the minimum error for the four circles are around 1.5–3. As expected, both of the two extreme conditions cannot lead the weighted average method to attain the best results.

(3) The root mean squared error (RMSE) and the hit rate

Table 5 shows the evaluation of the best prediction of the quantitative comparative approach, using the optimal weights of the characteristics variables and effect radius. The results display that the average of the 20% error hit rate is 78.0%. An interesting finding is that the ratios of the RMSE divided by the average house price in the same circle of supply and demand is rather stable, from 0.156 to 0.223.

(4) Spatial distribution of the actual and forecast house price

To show the spatial distribution of the actual and forecast house price per unit area, the results of Sindian are listed in Figs. 10 and 11. As expected, some hot areas around the MRT stations show high house price per unit area over 400,000 NT dollar per unit area. The north-east area is closer to the downtown of Taipei city, and its house price per unit area is much higher than the south-west area. The south is mountainous country and no transaction data. Comparing these two figures, the predictive model is rather accurate.

5.3. The results of conventional multivariate regression analysis

For comparison with the Quantitative Comparative Approach, we used the above same training data sets and testing data sets, and two classical Hedonic price approaches, multivariate regression analysis and neural networks, to establish the valuation models.

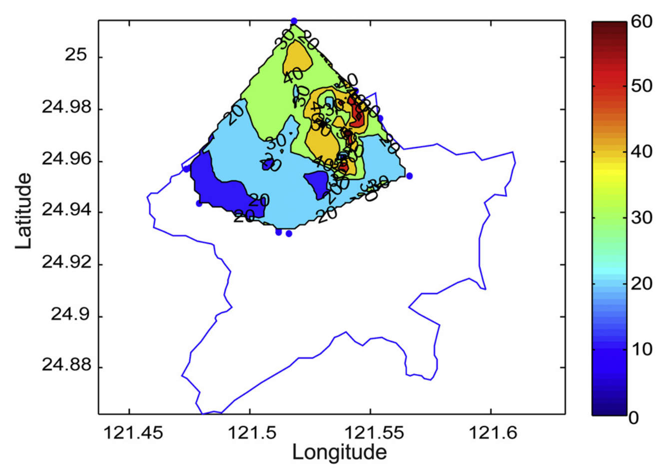

Fig. 10. The actual house price per unit area of Sindian (contour unit × 10000 NT dollar).

The testing data sets were employed to evaluate the accuracy of the results of the two classical Hedonic price approaches.

(1) Multivariate regression analysis

The following formula is employed to build the valuation models

$$Y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5N + a_6E + a_7N^2 + a_8E^2 + a_9N \cdot E \quad (36)$$

Table 6 shows the evaluation of the prediction of the multivariate regression analysis. The results display that the average of the 20% error hit rate is 68.7%. The ratios of the RMSE divided by the average house price in the same circle of supply and demand is rather stable, from 0.191 to 0.254.

Table 6
The prediction of the multivariate regression models.

circle of supply & demand	20% error hit rate	10% error hit rate	R^2	RMSE	Avg. house price	RMSE/Avg. house price
Sindian	74.3	41.8	0.530	8.65	37.98	0.228
Danshuei	64.7	37.2	0.387	4.71	21.27	0.222
Wunshan	72.0	42.3	0.296	8.34	43.68	0.191
Beitou	63.8	35.0	0.353	10.47	41.20	0.254
Average	68.7	39.1	0.392	8.04	36.03	0.224

Table 7
The prediction of the neural network models.

circle of supply & demand	20% error hit rate	10% error hit rate	R^2	RMSE	Avg. house price	RMSE/Avg. house price
Sindian	78.0	48.8	0.627	8.06	37.98	0.212
Danshuei	80.0	48.7	0.607	3.76	21.27	0.177
Wunshan	80.1	48.5	0.464	7.22	43.68	0.165
Beitou	66.4	37.7	0.467	9.42	41.20	0.229
Average	76.1	45.9	0.541	7.12	36.03	0.200

Table 8
The advantages and disadvantages of the real estate appraisal methods.

Approach	Advantage	Disadvantage
Income approach	• Sound financial theoretical base. • Don't need similar cases.	• High appraisal cost. • Time-consuming appraisal process. • Its estimated price is frequently deviate from the market price. • It cannot provide real comparative cases for appraisers to valuate real estate price.
Cost approach	• Sound financial theoretical base. • Don't need similar cases.	• High appraisal cost. • Time-consuming appraisal process. • Its estimated price does not consider marketability and profitability. • It cannot provide real comparative cases for appraisers to valuate real estate price.
Comparative approach	• It can provide real comparative cases for appraisers to valuate real estate price. In practice, comparative cases from real market trading are very important to persuade traders to accept the appraisal price. • It can handle the factors difficult to quantize. • Its estimated price is frequently close to the market price. Hence, it is generally accepted by appraisers.	• The correction coefficients often rely on the appraisers' subjective evaluation. The differences between appraisers and their subjectivity often make valuation results inconsistent and bring up questions. • Need a database consisting of many trading cases.
Hedonic price approach (Regression analysis)	• It does not have the human bias which often happens in the conventional comparative approach. • Low appraisal cost. • It is often used by the industry to deal with the mass appraisal of real estate.	• It cannot provide real comparative cases for appraisers to valuate real estate price. • It cannot handle the factors difficult to quantize. • Need a database consisting of many trading cases.
Quantitative Comparative Approach	• It can provide real comparative cases for appraisers to valuate real estate price. • It does not have the human bias which often happens in the conventional comparative approach. • Low appraisal cost.	• It cannot handle the factors difficult to quantize. • Need a database consisting of many trading cases.

(2) Neural networks

The multi-layered perception (MLP) is employed to build the valuation models. After a number of trials, the best network architecture and parameters which minimize the RMS error of testing data were selected as follows: number of hidden layer = 1, number of hidden unit = 5, learning rate = 0.6, number of learning cycle = 1500. Table 7 shows the evaluation of the prediction of the multivariate regression analysis. The results display that the average of the 20% error hit rate is 76.1%. The ratios of the RMSE divided by the average house price in the same circle of supply and demand is rather stable, from 0.165 to 0.229. The results are better than those of multivariate regression analysis but worse than those of the quantitative comparative approach.

5.4. Advantage and disadvantage of the real estate appraisal methods

The advantages and disadvantages of the real estate appraisal approaches are listed in Table 8. The main advantage of conventional comparative approach is that it can provide real comparative cases for appraisers to valuate real estate price. In practice, comparative cases from real market trading are very important to persuade traders to accept the appraisal price. The main disadvantage of conventional comparative approach is that the correction coefficients often rely on the appraisers' subjective evaluation. Our approach can keep the main advantage and avoid the main disadvantage of conventional comparative approach.

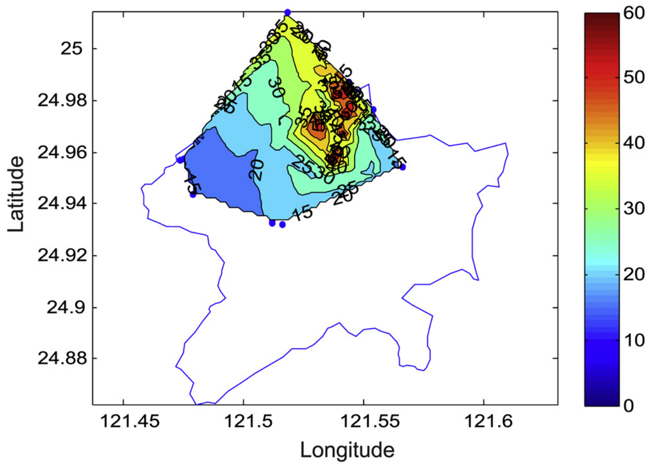

Fig. 11. The forecast house price per unit area of Sindian (contour unit × 10000 NT dollar).

6. Conclusions

The purpose of this paper is to propose an innovative real estate valuation approach which can calculate correction coefficient, so that it overcomes the shortages of subjective correction coefficient at the traditional comparison approach.

The independent variables include the distance to the nearest MRT station which represents the transportation function to the MRT station, the number of convenience stores in the living circle on foot which represents the living function in the living circle on foot, the age of house which represents the living function in room, the transaction date which represents the market trend, and the geographic coordinates which represent the transportation function to the downtown area.

The findings of the empirical results are as follows

(1) Using the sorting quantile method, the factors ranking by their importance from high to low are listed as follows: Distance to the nearest MRT station, Number of convenience stores in the living circle on foot, House age, and Transaction date.
(2) Overall, the distance of the nearest MRT station is the most important factor and its adjustment coefficient is inversely proportional to the distance of the nearest MRT station as expectation.
(3) The root mean squared error of the quantitative comparative approach is lower than that of the two classical Hedonic price approaches, multivariate regression analysis and neural networks. The 20% and 10% error hit rates are significantly higher than those of them. Therefore, the quantitative comparative approach is a promising approach for mass appraisal of real estate.

The suggestions of future research are as following.

(1) Firstly, the sequence of decomposition of factors in this study, no matter in which area, are the same. Maybe the sequence affects the prediction of the approach. Therefore, optimizing the sequence may further improve the prediction.
(2) Secondly, one advantage of the stepwise decomposition regression analysis is that it can build dimensionless adjustment coefficient models. These dimensionless models may have some generality across various circles of supply and demand. For example, the dimensionless adjustment coefficient model of the distance to the nearest MRT station built with the data set collected from the A circle of supply and demand may be suitable to the B circle. This hypothesis is worth further studying.
(3) Lastly, using the quantitative comparative approach, the expected error of the estimated price of the target case that can find many similar comparative cases in the same circle of supply and demand may be smaller than that of the target case that cannot find many similar comparative cases. If the hypothesis is true, a more reliable estimation of the expected error can be obtained.

References

[1] W.J. McCluskey, M. McCord, P.T. Davis, M. Haran, D. McIlhatton, Prediction accuracy in mass appraisal: a comparison of modern approaches, J. Property Res. 30 (4) (2013) 239–265.
[2] S. Peterson, A.B. Flanagan, Neural network hedonic pricing models in mass real estate appraisal, J. Real Estate Res. 31 (2) (2009) 147–164.
[3] W. McCluskey, P. Davis, M. Haran, M. McCord, D. McIlhatton, The potential of artificial neural networks in mass appraisal: the case revisited, J. Financ. Manage. Property Constr. 17 (3) (2012) 274–292.
[4] E.A. Antipov, E.B. Pokryshevskaya, Mass appraisal of residential apartments: an application of random forest for valuation and a CART-based approach for model diagnostics, Expert Syst. Appl. 39 (2) (2012) 1772–1778.
[5] C.C. Lin, S.B. Mohan, Effectiveness comparison of the residential property mass appraisal methodologies in the USA, Int. J. Hous. Markets Anal. 4 (3) (2011) 224–243.
[6] J. Zurada, A.S. Levitan, J. Guan, A comparison of regression and artificial intelligence methods in a mass appraisal context, J. Real Estate Res. 33 (3) (2011) 349–387.
[7] D. Friesen, M. Patterson, B. Harmel, A Comparison of multiple regression and neural networks for forecasting real estate values, Reg. Bus. Rev. 30 (2011) 114–136.
[8] V. Kontrimas, A. Verikas, The mass appraisal of the real estate by computational intelligence, Appl. Soft Comput. 11 (1) (2011) 443–448.
[9] N. Nghiep, C. Al, Predicting housing value A comparison of multiple regression analysis and artificial neural networks, J. Real Estate Res. 22 (3) (2001) 313–336.
[10] M. Al-Akhras, An evolutionary-optimised artificial neural network approach for automatic appraisal of Jordanian lands and real properties, Proceedings of the 11th WSEAS International Conference on Neural Networks (2010) 203–208.
[11] J. Kilpatrick, Expert systems and mass appraisal, J. Property Investment Finance 29 (4/5) (2011) 529–550.
[12] A.J. Gonzalez, R. Laureano-Ortiz, A case-based reasoning approach to real estate property appraisal, Expert Syst. Appl. 4 (2) (1992) 229–246.
[13] A.G. Musa, O. Daramola, E.A. Owoloko, O.O. Olugbara, A neural-CBR system for real property valuation, J. Emerging Trends Comput. Inf. Sci. 4 (8) (2013) 611–622.
[14] A. Mimis, A. Rovolis, M. Stamou, Property valuation with artificial neural network: the case of Athens, J. Property Res. 30 (2) (2013) 128–143.
[15] I.H. Gerek, House selling price assessment using two different adaptive neuro-fuzzy techniques, Autom. Constr. 41 (2014) 33–39.
[16] X.S. Liu, Z. Deng, T.L. Wang, Real estate appraisal system based on GIS and BP neural network, Trans. Nonferrous Metals Soc. China 21 (2011) 626–630.
[17] P.P. Bonissone, W. Cheetham, Financial applications of fuzzy case-based reasoning to residential property valuation, Barcelona, Spain, in: Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, vol. 1, 1997, pp. 37–44.