Prediction Model of House Price in Chiang Mai Province

**Damrongsak Rinchumphu[1,2], Pokpatapee Buosont[1,2], Nopnawat Lualai[1,2]
Setthawut Ketklin[1,2], Warut Timprae[2], Sarote Tepweerakun[2], Choo Yit Yang[3],
Non Phichetkunbodee[1,2]***

[1] Department of Civil Engineering, Faculty of Engineering, Chiang Mai University 50000, THAILAND
[2] Civil Innovation and City Innovation Research Laboratory, Chiang Mai 50000, THAILAND
[3] Shin Yang Composite Sdn Bhd, Sarawak 98000, MALAYSIA

* Corresponding author e-mail: nnonn.spirit@gmail.com
Received: 17 Nov 2020; Revised from: 8 Dec 2020; Accepted: 11 Dec 2020
Print-ISSN: 2228-9135, Electronic-ISSN: 2258-9194

Abstract

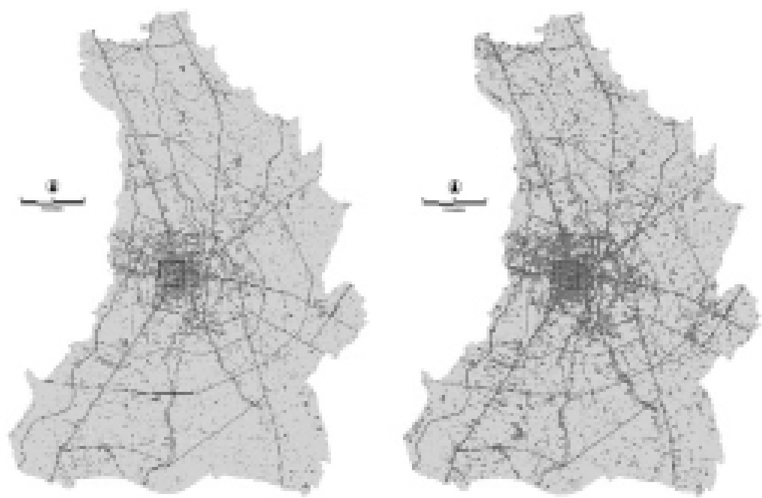
Pre-sale house price estimation has been a challenge for real estate developers, especially for emerging markets such as Chiang Mai. Therefore, this study calibrated and established a Hedonic Pricing Model to forecast the sales price of new houses from 125 residential development projects located in Chiang Mai, Thailand. In this study, a total of 22 variables were recorded, and the data were classified into 3 categories, which were location, physical attribute of the project and the nature of the house and land. The results showed 8 out of 22 variables were well incorporated in the model, and the Semi-Natural Logarithm form was determined as the most suitable model for predicting the pre-sale price at this target location. The asking price of a new detached housing project was found at 25.36% higher than other types of houses. Meanwhile, the newly-constructed house price was decreased by 9.98% if the project location was adjacent to or further away from the ring road, which was distant away from the Chiang Mai Metropolitan area. Furthermore, the house's pre-sale price was dropped when the available number of units in the project was large. The outcomes of this study indicate the effectiveness of the developed model as a proficient tool for estimating the competitive price for newly built properties in Chiang Mai.

Keywords: Prediction Model, House Price, Hedonic Price Model, Regression Analysis, Chiang Mai

1. Background

Shelter, besides food, clothing, and medication, is one of the four crucial requisites for human. In General, people seek for the most secure and beneficial as well as safest habitat. Social and economic status along with cultural traditions, comfort or other fundamental necessities, living conditions as well as the influence of various technological developments, have variously determined how humans select their house. Additionally, individual housing need is also differed according to socioeconomic status. Therefore, the decision of house purchasing is also relied on other factors aside prices. (Lorlertsakul & Sakworawich, 2017).

At present, newly developed residential projects in many cities, including Chiang Mai, have been extended outside their Metropolitan area due to population's growth. Srivanit (2010) compared the residential locations in Chiang Mai between 2000 and 2006 as present in Figure 1 (a) and Figure (b), which the expansion of the residential areas has been clearly evidenced (Srivanit, 2010). High demand of housing in Chiang Mai has attracted current and future real estate investments for promising opportunities. Due to high competitions, however, normal pricing may result in financial risks. Frequently, the pricing trend of new properties in the Chiang Mai urban areas has increased, while has decreased in the suburb areas (McGrath, Sangawongse, Thaikatoo, & Corte, 2017). Specifically, the city layout of Chiang Mai with the ring road located around the downtown provides the access for people commuting from outside the city. The new housing projects in Chiang Mai, therefore, has incorporated the distance from the Metropolitan area when the sales price is evaluated (Guntawilai & Sirisrisakulchai, 2017). However, this relationship is not yet certain, and location is not the sole factor for setting the target sales price of new residential properties. In addition to distances from the city, physical attributes of the project, the distance from the key authorities, the competitive projects, and the economic potential of the surrounding areas must be integrated in the pricing analysis of newly constructed houses (Malaitham, Fukuda, Vichiensan, & Wasuntarasook, 2020).



(a) (b)

Figure 1 Expansion trends of Chiang Mai – a) Year 2000 and b) Year 2006

There are several existing methods to develop a model analyzing appropriate sales pricing of new houses (Srivanit, 2010). One of the most accurate and widely adopted approaches of real estate-related datasets is Regression Analysis, which can be subdivided into many types (Rinchumphu, Kridakorn Na Ayutthaya, & Yunus, 2020). According to the housing study in Singapore, a spatial model was applied to analyze both sales and rent price of houses in all housing categories, including detached house, twin house, condominium, etc. (Lehner, 2011). Data testing in the same study for both the private and governmental sectors (Housing Development Board) was conducted in three comparative models: the least square approach model, the spatial error model, and the spatial durbin model. The results of Lehner (2011)'s study showed that the spatial error model was the most effective in Singapore. The space available per floor of the building and the distance from the central business district (CBD) were identified as remarkable influences on housing and rent prices at this study site. In Thailand, condominium pricing factors or attributes in Bangkok were classified into 3 aspects in accordance with the concept of the Hedonic price model: location attributes, structural attributes, and neighborhood attributes (Bunjam & Nilbai, 2017). Moreover, an econometric model was created for predicting the sales price per square meter, a dependent variable, for 141 condominium models when 21 independent variables were incorporated in the study. Bunjam & Nilbai (2017) sampled data from 40 projects, and employed 4 Regression Analysis models for data analysis. The model was constructed with specific forecasting variables respected to fundamental attributes of the condominium (5 variables) and a predictive model with the basic and specific attribute variables of the condominium (21 variables) by both models using straight line equations and the Log-Linear Equation. Bunjam & Nilbai (2017)'s results showed that 4 of 9 variables displaying statistical significance were the variables related to the basic attributes of the condominium, consisting of the inner Bangkok area, the distance from the Skytrain station, the condominium floor position, and the central maintenance fee. In addition, the 5 insignificant variables associated to condominium attributes were the furniture decoration of common area, the number of condominium floors, number of bedrooms, project area, and the readiness.

While Bunjam & Nilbai (2017)'s study exclusively focuses on the development of predictive models for the condominium in the Bangkok area where the population density was high, and the urban development model was vertical (Tochaiwat & Likitanupak, 2017), there is no similar studies available for residential projects in Chiang Mai. Chiang Mai differs compared to Bangkok in several aspects such as way of life, society, culture, and economy. Therefore, to analyze modelled sales pricing of Chiang Mai properties, the relationship of specific factors need to be considered in order to be explicitly consistent with the Chiang Mai residential development model. Our study, therefore, calibrated, tested and fully developed a forecast model for housing development projects in Chiang Mai, which can be implemented as an appropriate tool, which resulted in fair and competitive prices for recently built residential properties in Chiang Mai.

2. Research method

Data from Geographic Information System (GIS) and housing projects in Chiang Mai were collected in order to create a regression analysis model for sales price forecasting and calibrating as well as revising the model for achieving the most harmonious results toward model finalizing. There were 3 stages of the study process: information search, variable identification and data analysis, respectively, as follows.

2.1 Data Collecting

According to the available publications on the development of a sale-pricing forecast model of housing projects, the statistically significant variables observed and integrated to forecast housing prices in each study were different (Bunjam & Nilbai, 2017; Guntawilai & Sirisrisakulchai, 2017; Lehner, 2011; Tochaiwat & Likitanupak, 2017). The variables in this study were classified into 3 groups, which were project location, project physical attributes, and house as well as land attributes. The data were collected from 2 sources as described below.

1. Primary Data: Data were obtained from the survey of the area from direct sources such as location, distance from main streets, distances from key sectors, land area, number of buildings, etc.

2. Secondary Data: Data were gathered from researching project documents, advertisements, journals, related studies, and online information from various related sectors such as usable areas, number of bathrooms and bedrooms, number of other utility units, types of buildings and structures, etc.

2.2 Variable Identification

The sales price of houses located in Chiang Mai development projects was the dependent variable of this study, which the price of the plot was non-angle, in a blind spot or a special location of the projects, was good representative, and reflected the target market of each project appropriately. Information from a total of 242 data were included in this study analysis, which also includes the 2020 launched residential development. 22 variables were then categorized into 3 main groups as follows.

1. Project Location (A)
2. Project Physical Attributes (B)
3. Other Utilities (C)

The details of each variable group can be summarized as shown in Table 1.

Variable Group	No.	Detail	Symbol	Unit	Expected Direction
Project Location	1	Located on the ring road	A1		-
	2	Number of Unit	A2	Unit	-
	3	Distance from main road	A3	km	-
	4	Distance from the nearest hospital	A4	km	-
	5	Distance from the nearest school	A5	km	-
	6	Distance from the nearest shopping mall	A6	km	-
Project Physical Attributes	7	Town Home	B1		N/A
	8	Detached House	B2		N/A
	9	Semi-Detached House	B3		N/A
	10	Land Area	B4	Sq.W.	+
	11	Usable Area	B5	Sq.M.	+
	12	Number of Floor	B6	Floor	+
	13	Bedroom	B7	Room	+
	14	Restroom	B8	Room	+
	15	Parking Lot	B9		+
Other Utilities	16	Gym (Fitness)	C1		+
	17	SECURITY GUARD	C2		+
	18	CCTV	C3		+
	19	Park	C4		+
	20	Swimming Pool	C5		+
	21	Automatic door	C6		+
	22	Playground	C7		+

Table 1. Details of the variables obtained from searching for analysis.

According to Table 1, the factors based on the symbols of changing directions compared to the variables could be further classified into 2 subcategories as follows:

1. Variables with a positive correlation with sales price, were variables that, if the variables were increased or decreased, the housing prices would be influenced in the same direction. These variables included A1 – A3 / B7 – B12 / C1 – C3. Generally, when determining the correlation coefficients, these variables were marked with directional coefficients corresponding to or the same as the price variables.

2. Variables with an indeterminate direction with sales price, were variables that were changed, increased or decreased, would have no exact correlation with the housing price changes. These variables were B1 – B3. In general, the directional signs of these variables consisted of indeterminate coefficients, most of which were in the same direction as price variables, but sometimes the opposite direction.

After the variable identification, the next step was data analysis by choosing a statistical analysis tool, the Hedonic Price Model, which is discussed in the next section.

2.3 Data Analysis

For Hedonic Price Model simulation, regression analysis is one of the statistical methods that have been used to study the relationship between two or more variables (Tochaiwat & Likitanupak, 2017). The analysis aimed to forecast a Dependent Variable, sales price of houses in residential development project, from other independent variables that influenced the dependent variable. In this research, all variables listed in Table 1. The method is widely used in business, economics, agriculture, social sciences, and science. When there are more than one independent variables, the regression analysis is called multiple regression analysis. From retrieved data, a multiple regression analysis is the relationship between independent and dependent variables in a linear pattern (Tochaiwat & Likitanupak, 2017).

The nature of the relationship is formatted as Equation (1)

$$Y_i = B_0 + B_1X_{i1} + B_2X_{i2} + B_3X_{i3} + \cdots + B_kX_{ik} + \varepsilon_i \qquad \text{(Eq.1)}$$

The dependent variable Y (the sale price of the housing project) was varied according to the 22 independent variables X as shown in Table 1. The weight coefficient for each variable is B as the multiplier of each independent variable. The equation follows the agreement (Assumption) in regression analysis as follows:

1. The independent variable (X) and the dependent variable (Y) must be quantitative variables or Continuous Variables, or has a measurement level as Interval or Ratio Scale, such as weight, height, blood pressure, Cholesterol level, income, age, score, etc. If some independent variables (X) have a nominal or ordinal scale, the data must be converted to dummy variables, which the values equal to 0 and 1, before the analysis. In addition, the dummy variables should not be multiple because this will create more discrepancies.
2. Each independent variable has a linear relationship with the dependent variable.
3. Independent variables should not be related or independent (Correlation should not exceed 0.7) in the case of multiple regression analysis because it will cause Multicollinearity. In other words, if an independent variable is highly correlated, it will have an inflated effect on the decision coefficient (R square).
4. The distribution of the dependent variable is a normal distribution at all X values.
5. The variance of Y has the same variance for all X values.
6. The variance of the forecasting error (Residual) at all points on the regression line is the same.

3. Results

Multiple regression model was first generated from data from 242 Chiang Mai residential selling house data prior to developing a model estimating the sales price of new houses until the most statistically accurate model was determined. The model performance analysis, after analyzing the correlation with Pearson's Correlation between the 22 variables, indicated no relations with > 0.75, which suggested all variables could be included for model simulation (Rinchumphu, Kridakorn Na Ayutthaya, & Yunus, 2020).

To measure accuracy of each model, a coefficient of determination, coefficient of Multiple Determination, popularly known as R-Squared, are required as a statistic used to measure how equated this mathematical model is with data. The analysis revealed the highest R-Squared value was 0.741; in other words, the accuracy of the model was 74.1% of the data, which was more than other models. Therefore, in this study, this model would be used for further model creation for forecasting sales price.

The next step was to create a forecasting model using all the variables from the base model. The results of the analysis of the base model consisted of the coefficient of each sub-variable in the form of a semi-natural base logarithm or B value as shown in Table 2.

After deciding on the models, the next step would be the Analysis of Variance (ANOVA) of the base model. There are only 8 variables passed the statistics test and used in the model as follows:

1. Project location variable
   - A1: Located on the ring road
   - A2: Number of Unit
   - A6: Distance from the nearest shopping mall
2. project physical attributes variable
   - B2: Detached house
   - B4: Land
   - B5: Usable area
   - B9: Parking lots
3. Other Utilities variable
   - C1: Gym (Fitness)

All variables are conducted for the Analysis of Variance: ANOVA, and it was found that the statistic F (F-test) was equal to 36.754 and the statistically significant level at 95% value was 0.05. The value was very small that it was close to 0, indicating that this model was statistically justified to some extent.

In Table 2, there was a two-way test of the sub-variables in the model, showing the results of the two-way analysis in column t and statistical significance level values in column Significant of all the sub-variables in this model, the statistical significance level was all very small, which suggested this results could be used to create a predictive model by calculating the coefficients in the form of relationships. Consequently, the forecasting equation was determined as follows:

$$\begin{aligned} \text{Ln (Price)} = \; & 0.603 - 0.027\text{*(distance from shopping mall)} \\ & - 0.105\text{*(round of ring road)} + 0.002\text{*(usable areas)} \\ & + 0.007\text{*(land)} + 0.185\text{*(parking lots)} \\ & + 0.226\text{* (detached house)} - 0.001\text{*(number of units)} \\ & + 0.167\text{*(gyms)} \qquad \text{(Eq.2)} \end{aligned}$$

Model	Unstandardized Coefficients (B)	t	Significant	Marginal (%)
(Constant)	0.603	3.475	0.001	
A1	-0.027	-2.743	0.007	-2.69
A2	-0.105	-2.734	0.007	-9.98
A6	-0.001	-2.703	0.008	-0.06
B2	0.226	3.097	0.003	25.36
B4	0.007	5.219	0.000	0.71
B5	0.002	3.602	0.000	0.23
B9	0.185	3.129	0.002	20.35
C1	0.167	2.866	0.005	18.12

Table 2. Coefficient analysis results from the model

From Equation 2, every variable had a coefficient associated with it. If the coefficient was high, then that variable affected more on the sales price of houses. On the other hand, if the coefficient was low, that variable had slight effects on the sales price. In addition, the positive and negative signs referred to the direct and inverse variance of the relationships. The value of the coefficient from all 8 variables could be described as follows:

1. The distance from the nearest department store was in kilometers, and the coefficient was -0.027 or -2.69% of the price change rate. It meant if the project was far from the mall by 1 kilometer, the price would decrease by 2.69%.

2. There were 4 ring roads with a coefficient of -0.105 or -9.98% of the price change rate. It meant if the number of ring road turns was increased by 1, the sales price would decrease by 9.98%.

3. Usable area was the area utilized within the building. The unit was square meters. The coefficient was +0.002 or 0.23% of the price change rate. It could be concluded that if there was an additional 1 square meter of the area, the sales price would increase by 0.23%.

4. Land area was the area of land measured in square wa (1 square wa = 4 square meters). The coefficient was +0.007 or 0.23% of the price change rate. It indicated that if there was an additional 1 square wa of land area, the selling price would increase 0.71%.

5. Parking lot was the number of parking spaces within the house. The coefficient was + 0.185 or 20.35% of the price change rate. It implied that if there was 1 additional parking lot, the sales price would increase by 20.35%.

6. Detached house was to answer whether the house was a single-detached house or not. The coefficient is + 0.226 or 25.36% of the price change rate, meaning that f it was a single-detached house, the sales price would increase by 25.36% compared to other types of house.

7. The number of units in the project was the number of units released for each project. The coefficient is -0.001 or -0.06 marginal percent change of the sales price change rate. It meant that if there is an increase in the number of units by 1, the sales price would be reduced by 0.06%.

8. Gym (Fitness) was to answer whether there was a gym in the project or not. The coefficient is + 0.167 or 18.12% of the price change rate. It meant if there was a gym in the project, the sales price would increase by 20.35% compared to the projects without gym.

4. Conclusion

Our results revealed that the general factors or attributes of the houses in the housing project affected the pricing. The coefficient estimated from the model could be used to calculate the sales price of a house with the form of equation as follow.

$$Ln(Price) = 0.603 - 0.027*(\text{distance from shopping mall}) - 0.105*(\text{round of ring road}) + 0.002*(\text{usable area}) + 0.007*(\text{land area}) + 0.185*(\text{parking lot}) + 0.226*(\text{detached house}) - 0.001*(\text{number of units in the project}) + 0.167*(\text{gym})$$

The variables with the greatest effect on the rate of change in sales price were house attributes. If it was a detached house, it would cost 25.36% more than other types of houses because people who wanted to live outside the city limits had a need for a detached house which contained more private space than other forms of houses. The second most influential factor was the number of parking spaces, which was based on the assumption that if a house had a lot of parking space, it would be a larger house with more usable space and land area. As a result, that house had a higher price. The sales price would increase by 20.35% for a house with 1 additional parking lot. In addition, another important factor was how far the project location from the ring road. From the analysis results, it was found that if the number of ring roads increased, the sales price of houses in the project would be lower. One additional ring's number of rounds resulted in a 9.98% drop in home sales prices because commuting from the city limits to the next ring road during rush hour would take a long time due to heavy traffic. The farther the house was, the more time required for traveling to the Metropolitan area of Chiang Mai. However, an interesting factor was the number of units on sale in the project. If there were a large number of units in the project, the sales price of the house would drop. If the number of units increased by 1, the sales price of the house would drop by 0.06% due to the distribution of the developers who focused on the sales volume rather than enhancing the level of the houses within the project. If there was a large share of sales, the sales price would be lower due to the limitations of land area and the nature of the houses. It was also found that in the upper-level house projects, the number of units was significantly smaller while the prices were much higher than those in general projects.

The results of this study could be used as an effective and useful reference for project developers who wanted to invest in housing projects around city areas of Chiang Mai as a basis for determining the initial sales price in the future.

5. Recommendations

1. For the analysis of statistical data in this study, the variables used in the analysis should be quantitative variables with normal distribution with measurement level as the unit because it would be more appropriate than dummy variables with a scale value of 0 or 1 by the reason that dummy variable measurement could not imply amount or quantity. For example, having gym in the project should be measured as the number of places rather than measuring with 0 or 1. In other words, if the researcher could distribute the variables in quantities, the values of those variables could be analyzed and explained more accurately.

2. The models obtained from this study was based on a sample of houses in Chiang Mai's housing estate projects, the sample of the analysis. The results from the analysis illustrated only a group of variables affecting housing prices in the study area. However, if the results of this analysis were to be used to determine factors in the area for other case studies, this model should be tested within that area as well, which must have location conditions and similar uses in order to know the factors of other areas where they are the same or different. This will be able to accurately determine the main variables that affect the sales price of housing estate projects.

References

Bunjam, C., & Nilbai, T. (2017). *An analysis of factors determining the housing estate's price using hedonic price method: Case study of Bang KhunThian district, Bangkok* Paper presented at the The 9th National Conference of Economists.

Guntawilai, B., & Sirisrisakulchai, J. (2017). *Evaluation of state property land value of number 1723 by hedonic pricing method, tambon Donkaew, Mae Rim, Chiang Mai*. Chiang Mai University,

Lehner, M. (2011). Modelling housing prices in Singapore applying spatial hedonic regression.

Lorlertsakul, M., & Sakworawich, A. (2017). Spatial regression model predicting condominium price per square meter in Bangkok. *Academic Journal of Architecture, 66*.

Malaitham, S., Fukuda, A., Vichiensan, V., & Wasuntarasook, V. (2020). Hedonic pricing model of assessed and market land values: A case study in Bangkok metropolitan area, Thailand. *Case Studies on Transport Policy, 8*(1), 153-162.

McGrath, B., Sangawongse, S., Thaikatoo, D., & Corte, M. B. (2017). The architecture of the metacity: land use change, patch dynamics and urban form in Chiang Mai, Thailand. *Urban Planning, 2*, 53+.

Rinchumphu, D., Kridakorn Na Ayutthaya, T., & Yunus, R. (2020). Property price attributable to subdivision neighbourhood designs: Hedonic pricing model approach in Bangkok metropolitan region, Thailand. *International Journal of Built Environment Sustainability, 7*(3), 37-47.

Srivanit, M. (2010). *Effects of climate change and spatial characteristics on urban heat island in Chiang Mai*. Paper presented at the The First Climate Thailand Conference 2010.
Tochaiwat, K., & Likitanupak, W. (2017). Bangkok four-star hotels' average daily rate (ADR) prediction model. *Pacific Rim Property Research Journal, 23*, 1-16. doi:10.1080/14445921.2017.1375663