Predicting Construction Cost Using Regression Techniques For Residential Building

Puguh Novi Prasetyono¹, Mas Suryanto HS¹, Hasan Dani¹

¹Department of Civil Engineering, Faculty of Engineering, Universitas Negeri Surabaya, 60231 Ketintang, Surabaya, East Java, Indonesia.

puguhprasetyono@unesa.ac.id

Abstract. Building prices fluctuate greatly each year. Price fluctuations can be on the prices of building materials, human resources, and other cost requirements used in construction. This fluctuation makes it difficult to estimate how much it will cost to build a house in the next year. There are several studies that have discussed budget plan predictions, but few estimate the planned budget for housing development, especially in Indonesia, because it is not yet known how to predict the cost of housing construction in the future, from here it is necessary to conduct research to obtain a formula in predicting development costs housing. The method is calculating the cost modeling using the regression method, the regression method used is Linear Regression. linear regression studies the linear relationship between the dependent variable Y and a single independent variable X. As a result, the formula obtained for the prediction of the cost of housing construction is \( y = 1,169,813.19 + (334,729,758)x \) where \( y \) is cost is predicted while \( x \) is what year with \( r^2 \) is 0.10.

Keywords: Regression, Residential Building, Construction Cost

1. Introduction
The process of calculating the budget plan is a step that is often taken in the implementation of building construction [1]. The level of accuracy in planning the cost budget is an important factor in determining the success of the project, with how much-estimated costs will be used, the owner will be able to know whether the project is feasible or not to be continued in its development from the financial side.

Performing the cost calculation stage for each definite estimator uses an index called the construction cost index [1]. The index is a number that shows the index for the price of wages, materials and materials. Of course, the index always changes every year which fluctuates in accordance with existing conditions in the market [2]. The price index is very useful for all stakeholders in calculating the cost budget plan, for example the index is very helpful for contractors when carrying out bids during tenders, then for the owner the index is very useful for determining the value of the building to be planned so the owner can find out how much it costs. will be issued in the development process [1]. The calculation of the budget plan implemented in Indonesia refers to the first two regulations, the HSPK (Main Activity Unit Price). The second one uses the Indonesian National Standard (SNI).

This budget plan is very helpful in the development process, one of the benefits is as a guide in controlling planning costs with the cost of implementing a construction project [3]. This budget plan is very old and used to be used in the implementation of construction projects [4]. Construction projects that use this cost budget plan include building dams, building bridges, building gedlung, etc.
For buildings, one of them is residential buildings. Calculation of the cost of residential buildings is often done because residential buildings are in addition to being a protection house, also used as an asset for future investment [5][6].

The cost of building a residential house is very dynamic and fluctuating [1]. The price fluctuation can be on the price of building materials, human resources, and other cost requirements used in construction. This economic uncertainty could have a major impact on the business especially on long-term projects [1]. To minimize the risk of uncertainty in investment costs for the construction of building a residential house, it is necessary to predict the cost of residential building construction.

In some countries, several studies have been conducted that discuss housing, especially those related to the prediction of budget plans. Existing research includes: use of multiple linear regression, neural networks, and case-based reasoning to predict budget costs based on historical cost data [7], New conceptual system for predicting the cost of housing construction in the Kingdom of Saudi Arabia (KSA) explains a brief model-the main models of the system developed to free continuity and focus on the energy cost prediction model of residential buildings in Saudi Arabia [8], residential construction cost An Italian survey [9], Residential outage cost estimation: Hong Kong [10]. Estimation and prediction of construction cost index using neural networks, time series, and regression [1]. There are also those who try to combine several cost predictions to detect early potential cost overrun of the Project [11], Regression methods have also been used in research to predict initial cost estimates using multivariate analysis and regression factors [12]. but only a few have predicted the planned budget for housing construction in Indonesia so this research is carried out.

2. Method

In the data collection process carried out, among others, are as follows, first is looking for supporting data used in the process of calculating the HSPK (Main Activity Unit Price) and Indonesian National Standard (SNI), then the work plan and requirements, as well as work drawings that will be used in development. The next step is to conduct interviews with stakeholders regarding missing data, such as the latest labor wages, the latest material prices, etc. because data such as material prices, labor wages are the initial data needed in calculating the estimated construction costs and without such data, the calculation of estimated building costs will be carried out. [13][14]. Then from the data that has been obtained, the next step is to estimate the cost using the quantity take-off method. The data calculated is data on the cost of housing construction in 15 years.

Based on the obtained budget plan data continued by calculating the cost modeling using the regression method, the regression method used is linear regression. In the use of linear regression is to determine the relationship between the dependent variable Y and the independent variable X so that the equation can be found [15]. This linear regression model is used to provide a dependent description of the variable in the straight line equation that is defined as follows:

\[ Y = a + bX \]  

- \( a \) is the y-intersect of the line,
- \( b \) is its slope.
- \( Y \) be the dependent variable
- \( X \) be the independent variable

The coefficient of determination is then defined as follows:

\[ r^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \]  

- \( n \) be the number of observations
• $\hat{y}_i$ be the estimated value of the dependent variable for the $i$th observation, as computed with the regression equation
• $y_i$ be the observed value of the dependent variable for the $i$th observation
• $\bar{y}$ be the mean of all $n$ observations of the dependent variable

$r^2$ is the fraction of the overall variance that is explained. The closer the regression model’s estimated values $\hat{y}_i$ lie to the observed values $y_i$, the nearer the coefficient of determination is to 1 and the more accurate the regression model is.

### 3 Results and Discussion

#### 3.1 Budget plan

In calculating the estimated cost of housing construction it is necessary to know the cost of housing construction each year for 15 years to then be compared. The following is a recapitulation of type 36 housing budget plans for the past 15 years as follows as Table 1:

| No. | Year    | Total price  | Price/m² |
|-----|---------|--------------|----------|
| 1   | 1st year | (IDR) 77,358,117,82 | (IDR) 2,149,000,00 |
| 2   | 2nd year | (IDR) 81,548,197,68 | (IDR) 2,265,000,00 |
| 3   | 3rd year | (IDR) 87,658,030,14 | (IDR) 2,436,000,00 |
| 4   | 4th year | (IDR) 93,659,464,05 | (IDR) 2,602,000,00 |
| 5   | 5th year | (IDR) 103,911,802,36 | (IDR) 2,886,000,00 |
| 6   | 6th year | (IDR) 113,775,174,10 | (IDR) 3,160,000,00 |
| 7   | 7th year | (IDR) 118,560,027,85 | (IDR) 3,293,000,00 |
| 8   | 8th year | (IDR) 122,195,864,96 | (IDR) 3,394,000,00 |
| 9   | 9th year | (IDR) 129,112,903,29 | (IDR) 3,586,000,00 |
| 10  | 10th year| (IDR) 129,855,698,17 | (IDR) 3,607,000,00 |
| 11  | 11th year| (IDR) 135,942,870,78 | (IDR) 3,776,000,00 |
| 12  | 12th year| (IDR) 180,243,728,08 | (IDR) 5,007,000,00 |
| 13  | 13th year| (IDR) 252,094,445,83 | (IDR) 7,003,000,00 |
| 14  | 14th year| (IDR) 266,771,084,49 | (IDR) 7,410,000,00 |
| 15  | 15th year| (IDR) 296,368,400,00 | (IDR) 8,323,000,00 |

The results of the calculation of the cost budget plan there is a fluctuating level of the cost of housing construction each year. The price fluctuation can be on the price of building materials, human resources and other cost requirements used in construction.

Based on the data budget plan, it can be seen that every year there is an increase in fluctuating costs, there is a tendency to increase the price of residential buildings every year. For more details, it can be presented in figure 1 below:
Figure 1 Increase in house price / m2 per year

The increase in house price / m2 was caused by many factors, these factors could be due to an increase in material costs or due to labor costs. Several years have experienced a fairly high increase between the years 12 to 13, after identifying the increase was due to a significant increase in wages. Wages from 12 to 13 almost doubled. In addition to other causes of wages is material price increases, price increases can be caused by rising material prices from producers. In the 12th and 13th years the increase in MSEs (City / District Minimum Wages) followed by rising fuel prices.

From the data that has been obtained about the annual house price, then a calculation is performed to identify or predict the estimated cost of housing construction in the following year, the method used is a linear regression method so that later the formulation will be obtained to predict the cost of housing construction.

3.2 Cost modeling using linear regression

This simple linear regression illustrates the relationship between two variables, first is the dependent variable which is represented by Y with the dependent variable which is represented by X. The analysis is used to obtain the relationship between the independent variable and the dependent variable where the relationship will show a positive relationship or relationship, negative, this aims to determine or predict the value of the dependent variable if there is an increase or decrease in the independent variable. To predict the budget plan, the type of regression used can be used, one of which is the linear regression method. This linear regression as previously explained that using two variables, namely the independent variable X and the dependent variable Y. of these two variables will be arranged into a linear equation which is described in a straight line with the equation $Y = a + bX$, for the value of Xi, is the year whereas for the value of Yi is the cost incurred to build a house in a certain year, while the value of an in the calculation of the value of a is 1169813.19 then the value of b is 334,729.76 of that number do the calculation of the value of y. The following is a Cost modeling using linear regression as follow as Table 2:

| No | Xi   | Yi      | Xi.Yi   | Xi   | Yi=a+bx  |
|----|------|---------|---------|------|----------|
| 1  | 1    | 2.149.000,00 | 2.149.000,00 | 1    | 1.515.000,00 |
| 2  | 2    | 2.265.000,00 | 4.530.000,00 | 4    | 1.859.703,30 |
| 3  | 3    | 2.436.000,00 | 7.308.000,00 | 9    | 2.204.406,59 |
| 4  | 4    | 2.602.000,00 | 10.408.000,00 | 16   | 2.549.109,89 |
| 5  | 5    | 2.886.000,00 | 14.430.000,00 | 25   | 2.893.813,19 |
| 6  | 6    | 3.160.000,00 | 18.960.000,00 | 36   | 3.238.516,48 |
| 7  | 7    | 3.293.000,00 | 23.051.000,00 | 49   | 3.583.219,78 |
| 8  | 8    | 3.394.000,00 | 27.152.000,00 | 64   | 3.927.923,08 |
| 9  | 9    | 3.586.000,00 | 32.274.000,00 | 81   | 4.272.626,37 |
| 10 | 10   | 3.607.000,00 | 36.070.000,00 | 100  | 4.167.329,67 |
| 11 | 11   | 3.776.000,00 | 41.536.000,00 | 121  | 4.962.032,97 |
| 12 | 12   | 5.007.000,00 | 60.084.000,00 | 144  | 5.306.736,26 |
| 13 | 13   | 7.003.000,00 | 91.039.000,00 | 169  | 5.651.439,56 |
| 14 | 14   | 7.400.000,00 | 103.600.000,00 | 196  | 599.028,57 |
| 15 | 15   | 52.564.000,00 | 472.591.000,00 | 1015 | 6.340.758,24 |

In determining the linear model, it is necessary to have the $r^2$ test or commonly called determination. Determination can be used to predict the contribution of the effect of variable x to
variable y. From the linear regression calculation in table 2, the value of $R^2$ is 0.1 with the regression formula obtained is $y$ obtained by the equation $a + bX$ or $y$ value of $1.169.813.19 + (334.729.758)X$, where the value of $X$ is the year of prediction.

4. Conclusion

After reviewing the results of calculations and discussion of Predicting Construction Cost Using Regression Techniques For Residential Building, it can be concluded as follows as the formula obtained for the prediction of the cost of housing construction is $y = 1,169,813.19 + (334,729,758)X$ where $y$ is cost is predicted while $X$ is what year with $r^2$ is 0.10.

References

[1] Y. Elfahham, “Estimation and Prediction of Construction Cost Index Using Neural Networks, Time Series, and Regression,” *Alexandria Eng. J.*, 2019.
[2] T. P. WILLIAMS, “Predicting Changes In Construction Indexes Using Neural Networks,” *J. Constr. Eng. Manag.*, vol. 120, no. 2, pp. 306–320, 1994.
[3] M. A. Momeni, S. Yaghoubi, M. Reza, and M. Aliha, “An Optimal Control Model for Analyzing Quality Investment in the Project Management,” *Comput. Ind. Eng.*, 2019.
[4] C. G. Wilmot, M. Asce, and B. Mei, “Neural Network Modeling of Highway Construction Costs,” no. July, pp. 765–771, 2005.
[5] A. Creach, E. Bastidas-Arteaga, S. Pardo, and D. Mercier, “Adaptation of Residential Buildings to Coastal Floods: Strategies, Costs and Efficiency,” *Clim. Adapt. Eng.*, pp. 245–270, 2019.
[6] F. Allon, *Home as investment*, vol. 2. Elsevier Ltd., 2012.
[7] G. H. Kim, S. H. An, and K. I. Kang, “Comparison of Construction Cost Estimating Models Based on Regression Analysis, Neural Networks, and Case-Based Reasoning,” *Build. Environ.*, vol. 39, no. 10, pp. 1235–1242, 2004.
[8] A. Alshibani and O. S. Alshamrani, “ANN/BIM-Based Model for Predicting the Energy Cost of Residential Buildings in Saudi Arabia,” *J. Taibah Univ. Sci.*, vol. 11, no. 6, pp. 1317–1329, 2017.
[9] R. Canesi and G. Marella, “Residential Construction Cost: An Italian Survey,” *Data Br.*, vol. 11, pp. 231–235, 2017.
[10] C. K. Woo, T. Ho, A. Shiu, Y. S. Cheng, I. Horowitz, and J. Wang, “Residential Outage Cost Estimation : Hong Kong,” vol. 72, no. 2014, pp. 204–210, 2020.
[11] B. Kim and Y. H. Kwak, “The Management of Operations Improving the accuracy and operational predictability of project cost forecasts: an adaptive combination approach,” *Prod. Plan. Control*, vol. 7287, no. May, pp. 1–18, 2018.
[12] S. M. Trost, M. Asce, G. D. Oberlender, and F. Asce, “Predicting Accuracy of Early Cost Estimates Using Factor Analysis and Multivariate Regression,” vol. 129, no. 2, pp. 198–204, 2003.
[13] J. S. Shane, K. R. Molenaar, S. Anderson, and C. Schexnayder, “Construction Project Cost Escalation Factors,” *J. Manag. Eng.*, vol. 25, no. 4, pp. 221–229, 2009.
[14] A. Touran and R. Lopez, “Modeling Cost Escalation in Large Infrastructure Projects,” *J. Constr. Eng. Manag.*, vol. 132, no. 8, pp. 853–860, 2006.
[15] R. Article, “Linear Regression Analysis,” vol. 107, no. 44, pp. 776–782, 2010.