Cost Estimation Process for Construction Residential Projects by Using Multifactor Linear Regression Technique

Abbas Mohammed Burhan Alshemosi¹, Humam Saad Hussein Alsaad²

¹, ²University of Baghdad, College of Engineering, Civil Engineering Department, Baghdad, Iraq

Abstract: Cost estimation in early stage plays an important role in the achievement of any construction project. Cost estimating is an assessment of the expected cost of any construction project. The accuracy of such estimate has a serious effect on the expected profit of the construction contractor. Hence, a certain contingency percentage should be added to the base estimate to improve the level of confidence. The process of cost estimating is materially affected by many factors. This paper aims to build a mathematical model for predicting the cost estimation for the residential construction projects. Multifactor linear regression technique is developed and used for predication of the cost of residential projects in Iraq. Ten effectiveness factors are used for cost estimation forecasting by (MLR) model, they include, experience of estimator, market status economic instability, accuracy of bidding documents, site condition and total area of residential building, the height of residential building, project location, type of residential building, estimating method. It was found that Multifactor linear regression has the flexibility to foretell the cost with a excellent degree of accuracy 91.76%, mean absolute percentage error 8.240% and coefficient of correlation (R) was 95.74%.

Keywords: Cost estimation, early estimate, accuracy of cost estimation, MLR Model, SPSS.

1. Introduction

Cost estimation is an essential part of construction residential projects, where cost is considered as one of the main criteria in decision making at the early stages the reliability of cost estimation has a significant effect on the success of construction residential projects [1].

For example, it may have a serious effect on contractor capability to compete successfully with other contractors. It also possess an important effect on contractor’s profit. Therefore, this research is an attempt to build and evaluate an analytical model which may be used to estimate the total cost of residential projects, by using Multifactor Linear Regression Technique (MLRT)[2].

2. Objective

The objective of this research to build and evaluate an analytical model which may be used to estimate the total cost of residential projects, by using Multifactor Linear Regression Technique (MLRT), through the following procedures:

1) Determination of Multifactor Linear Regression model variables, these variables having a high effect on the cost of residential projects in Iraq
2) Construct a comprehensive tool for residential projects cost estimation by using the Multifactor Linear Regression as the most efficient tool in predictions
3) Built an equation to find the cost estimation of construction residential projects.

3. Construction Cost Estimation

The cost estimation is a basic process for each project, it can be defined as the process of calculating the quantities of materials, labor, and equipment expected to be used to complete the work of the construction project within specific specifications, calculating the cost of the project shall be according to that [3].

3.1 The Goal of Estimation

The estimation is the expectations of the cost of implementing the work, where it is difficult to predict certain things such as worker productivity and working conditions. As many parties estimate the construction costs for different purposes. The predictability depends heavily on data from previous experience taking into account the differences between the new situation and past rates for realistic and real estimation [4].

4. Research Methodology

The research methodology followed up in carrying out the field investigation includes:

1) Literature survey
2) Preliminary interviews (Open questionnaire).
3) Building a questionnaire form.
4) Research sample
5) Distribution a questionnaire form
6) Displaying and discussing the results and drawing the final conclusions and recommendations.

Personal interviews were conducted by the researcher with the senior level of engineers working in the construction residential projects. The structured interview questionnaire started with ten questions. The questions which were divided into two parts, the first part includes general background information about the research sample, while the second part includes information about the cost estimation methods and factors affecting the process of cost estimate[4]. The interviews were conducted with a number of experts with great experience in the field of construction projects. Those experts involve engineers in Roads and Bridges Directorate, National Center for Engineering Consultancy, Engineering
Construction Office, Al-Faw General Engineering Company, Al-Mansour Contracting Company, professors in colleges of engineering in Iraq and engineers in private sector. All of them have more than twenty years working experience in the construction sector in Iraq. Moreover, the expert interviews were helpful in building the baseline questions that focusing on construction projects cost estimation.

Total number of (100) forms were distributed. Only (85) forms were completely returned.

The questionnaire forms were distributed to a number of the Iraqi state institutions as follows:

2) The Ministry of Planning
3) The engineering department in Baghdad governorate.
4) A number of academic professors in Engineering colleges (Baghdad University, AlNahrain University, AlMasour University College, and Al-Esraa University College).
5) A number of experts working in private sector companies (Almajd construction contracting company, Alhadbaa construction contracting company, and Almusawy construction contracting company).

5. Result Analysis

Keeping in mind the end goal to have the capacity to pick the correct technique for examination, the estimation level must be clear. For each kind of estimation, there are suitable methods that can be utilized. In the following research, ordinal scales were utilized. Ordinal scale as appeared underneath in Table (1) is a positioning framework that regularly utilizes numbers in climbing or diving request. The numbers allocated to the essential (10, 30, 50, 70,90) don't demonstrate that the interim between scales are equivalent, nor do they show supreme amounts [5-8].

With a specific end goal to evaluate information investigation and ascertain the Relative significance (RI), a theoretical Weight Value (WV) was joined to each of the five classes of answers availableThis Weight Value (WV) was compared to the mid-scope of each class on the five class decimal scale as clarify in Table (1)[9-11].

Relative Importance (RI): 
\[ RI = \frac{\sum (f_i \times W_i)}{N} \]  

\[ \text{Percentage} = \frac{\text{Number of frequencies}}{\text{total number of the answers}} \times 100 \]

\[ \text{Eq.(1)} \]

\[ \text{Eq.(2)} \]

Illustrates the methods of cost estimation results shown that the highest rate was for Expert judgment by (40%), and by (35%) for Square-meter estimates, and by (8%) for Parametric estimating and by (10%) for Cubic-meter estimates while only (7%) for Analogous estimates.

5. Factors that Affect Accuracy of Construction Cost Estimation

Based on the literature review and the results of interviews (open questionnaire), factors that influences the process of building cost estimation was determined. These factors were used as a base for a questionnaire survey. Such survey identifies the most important factors influencing on the accuracy of the cost estimation. These factors were as the display in Table (2).[2-18]

### Table 1: Weight Value of Descriptive Frequencies

<table>
<thead>
<tr>
<th>Category</th>
<th>Category Interval</th>
<th>Weight Value (WV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not important</td>
<td>0-20</td>
<td>10</td>
</tr>
<tr>
<td>Less important</td>
<td>20-40</td>
<td>30</td>
</tr>
<tr>
<td>somewhat important</td>
<td>40-60</td>
<td>50</td>
</tr>
<tr>
<td>Important</td>
<td>60-80</td>
<td>70</td>
</tr>
<tr>
<td>Very important</td>
<td>80-100</td>
<td>90</td>
</tr>
</tbody>
</table>

### Table 2: Factors that may affect accuracy of construction residential projects cost estimating

<table>
<thead>
<tr>
<th>No</th>
<th>Category</th>
<th>Factors affecting in the accuracy of cost estimation</th>
<th>Relative Importance%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Experience of cost estimator</td>
<td></td>
<td>87.65</td>
</tr>
<tr>
<td>2</td>
<td>Estimation method</td>
<td></td>
<td>80.59</td>
</tr>
<tr>
<td>3</td>
<td>The time allocated to estimation work</td>
<td></td>
<td>74</td>
</tr>
<tr>
<td>4</td>
<td>Updating cost information</td>
<td></td>
<td>65.76</td>
</tr>
<tr>
<td>5</td>
<td>Availability of the moderate cost indicators</td>
<td></td>
<td>78</td>
</tr>
<tr>
<td>6</td>
<td>Availability of productivity criterion</td>
<td></td>
<td>71.88</td>
</tr>
<tr>
<td>7</td>
<td>Procedure used in determining emergency and</td>
<td></td>
<td>44.24</td>
</tr>
</tbody>
</table>

Figure 1: Methods of cost estimation
6.1 Most Important Factors

By examining the importance indices of each factor in the table (3), it can be identified the factors that are heavily considered to have a high impact on the accuracy of cost estimate process. These factors, for example, the experience of an estimator, economic instability, accuracy of bidding documents, Site condition, the total area of the residential building. On the other hand, factors such as Social, cultural and ideological impact, the time between project announcement and bid opening average, conditions of weather, the number of elevators in the residential building, have low impact.

6.2 Pareto Analysis Technique (PAT)

Assumes that 20% of factors can have the most important effect. Therefore, this percentage will lead us to the selection of the first ten factors as shown in table (3).[13]

Table 3: Factors that may affect accuracy of construction residential projects cost estimating by using Pareto Analysis Technique (PAT)

<table>
<thead>
<tr>
<th>No</th>
<th>Factors affecting in the accuracy of cost estimation</th>
<th>RII</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>experience of estimator</td>
<td>87.65</td>
</tr>
<tr>
<td>2</td>
<td>Market status</td>
<td>86.94</td>
</tr>
<tr>
<td>3</td>
<td>Economic instability</td>
<td>86.47</td>
</tr>
<tr>
<td>4</td>
<td>Accuracy of bidding documents provided by client</td>
<td>85.29</td>
</tr>
<tr>
<td>5</td>
<td>Site condition</td>
<td>84.82</td>
</tr>
<tr>
<td>6</td>
<td>Total area of residential building</td>
<td>83.88</td>
</tr>
<tr>
<td>7</td>
<td>The height of residential building</td>
<td>83.41</td>
</tr>
<tr>
<td>8</td>
<td>Project location</td>
<td>81.76</td>
</tr>
<tr>
<td>9</td>
<td>Type of residential building</td>
<td>81.29</td>
</tr>
<tr>
<td>10</td>
<td>Estimating method</td>
<td>80.59</td>
</tr>
</tbody>
</table>

Where: RII which mean relative important index

These variables are represented fell into two classifications, subjective and objective variables. All variable can be analyzed so as to find the optimum approach to illustration for the variable within the modeling process.

Objective variables content three variables and can be measured using the unit of measurements such as the first variable experience of the estimator (x1) measured in a number of estimator years and the sixth variable total area of residential building (x6) measured in a square meter. While the seventh variable Height of residential building (x7) measured in the meter.

Subjective variables content seven variables and may be measured relying on coding system, as an example, the second variable Market status(x2) can be classified to stable and medium stable and unstable and assigns them the value 1, 2 and 3 respectively. Also, the third variable Economic instability(x3) can be classified to stable and unstable. It assigns them the values of 1, 2 and 3 respectively. Latterly, the ten variable Estimating method (x3) defines a method that used in Iraqi residential construction projects to cost estimation in early stage with five types Parametric estimate, Square-meter cost estimate, Estimates by the expert consultation, Cubic-meter cost estimate, Cost estimates for each user. It assigns them the values of 1,2,3,4,and 5 respectively. The forth variable the Accuracy of bidding documents can be classified to accurate, medium accurate and not, accurate. It assigns them the values of 1, 2 and 3 respectively. The fifth variable the Site condition can be classified to good condition, medium condition and poor condition, it assigns them the value 1,2 and 3 respectively. The ninth variable Type of residential building can be classified to suitable and not suitable, It assigns them the values of 1, 2 respectively. The eighth variable The project location can be classified to suitable and not suitable, It assigns them the values of 1, 2 respectively.

7. Development of Multiple Linear Regression (MLR) Model

A lot of many subscriptions and books and thesis explained the (MLR). In this paper, multifactor linear regression technique can use the optimum forecasting model and an extremely powerful statistical tool that permits the researchers to find relationship between dependent variables (actual cost) and independent variables (objective and subjective variables) (e.g. estimating method, experience of estimator etc.) It can tentatively formulate:

\[ Y = A0 + A1 \times x1 + A2 \times x2 + \ldots + C \text{Eq (1)} \]

Where:
1) \( x_i \) and \( x_j \) are the values of the input variables for the \( i \)th experimental run
2) \( Y \) is the corresponding response
3) C is the blunder segments that are thought to be free
typical factors with mean zero and fluctuation σ2
4) A0, A1, the regression parameters, [13]

The developers of the Statistical Package for the Social
Sciences let all work to do the product easy to utilize.
Additionally, the it has been trusted in this research that
SPSS is one of the prescient investigation programs, can be
anticipated with unquestionably what will occur next and can
be settled on more quick witted choices. In this study, it was
anticipated with unquestionably what will occur next and can
be settled on more quick witted choices. In this study, it was
utilized SPSS to headway
be settled on more quick witted choices. In this study, it was
anticipated with unquestionably what will occur next and can
be settled on more quick witted choices. In this study, it was

Relapse examination produces a condition to portray the
measurable connection between at least one indicator factors
and the reaction variable. SPSS programming marks the
institutionalized relapse coefficients as "Beta", whereas the
unstandardized coefficients are named "B". It has been
trusted a relapse forms as a capable device that could be
utilized to anticipate the charge of the private venture from
multi-factors, and it permitting to expectation the cost of
undertakings in future relying upon the data about past or
introduce ventures. Estimations of unstandardized
coefficients in a table (5.5) could be utilized to manufacture
the multifactor direct relapse, as demonstrated the
accompanying condition:

\[ \text{Total estimation cost} = 1.299 + 0.321X1 - 0.321X2 + 0.155x3 - 0.015X4 + 0.117X5 + 0.876X6 - 0.047X7 + 0.213X8 - 0.265X9 - 0.316X10 \]

8. Validity of the Multiple Linear Replace
(MLR)

The objective behind this the validation can be to realize a
measurable dependable form. Through the current research,
Ten novel papers for construction residential ventures. These
types of papers hadn't been integrated into form that was
made, other than applied for validation and confirmation of
the following form. The real charge of these ventures besides
the forecasted charge as shown in the table (7).

\[ MPE = \frac{\sum |E - A|}{n} \times 100 \]

\[ = \frac{0.824}{10} \times 100 \approx 8.24 \]

\[ R^2 = 93.54 \]

Table 4: Shows the summary of statistical analysis.

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>Range</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>20</td>
<td>1</td>
<td>4.98287</td>
</tr>
<tr>
<td>X2</td>
<td>20</td>
<td>1</td>
<td>1.08094</td>
</tr>
<tr>
<td>X3</td>
<td>20</td>
<td>1</td>
<td>0.50262</td>
</tr>
<tr>
<td>X4</td>
<td>20</td>
<td>1</td>
<td>1.29371</td>
</tr>
<tr>
<td>X5</td>
<td>20</td>
<td>1</td>
<td>1.48936</td>
</tr>
<tr>
<td>X6</td>
<td>20</td>
<td>5.08</td>
<td>0.5280</td>
</tr>
<tr>
<td>X7</td>
<td>20</td>
<td>9</td>
<td>1.97084</td>
</tr>
<tr>
<td>X8</td>
<td>20</td>
<td>1</td>
<td>0.51299</td>
</tr>
<tr>
<td>X9</td>
<td>20</td>
<td>1</td>
<td>0.51042</td>
</tr>
<tr>
<td>X10</td>
<td>20</td>
<td>1</td>
<td>0.51299</td>
</tr>
</tbody>
</table>

Table 5: Results of the correlation

<table>
<thead>
<tr>
<th></th>
<th>Actual cost model 1</th>
<th>Model Fit statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>98.0</td>
<td>R-squared 96.1 Adjusted R-squared 91.7</td>
</tr>
</tbody>
</table>

Where: The quantity (R), called the linear correlation
coefficient, which is used to measure the power of the
connection linking two factors or more. And the coefficient
determination (R2) which is used to know the explanatory
power of the model estimated (The estimated equation).And
Adjusted R-squared which is used to interpret the
explanatory power of the multiple linear regression models It
is derived from (R2).

Table 6: Regression Coefficients Values

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>B</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>1.299</td>
<td>1.385</td>
<td></td>
</tr>
<tr>
<td>X1</td>
<td>0.321</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>X2</td>
<td>-0.321</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>X3</td>
<td>0.155</td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td>X4</td>
<td>-0.015</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>X5</td>
<td>0.117</td>
<td>0.022</td>
<td></td>
</tr>
<tr>
<td>X6</td>
<td>0.876</td>
<td>0.248</td>
<td></td>
</tr>
<tr>
<td>X7</td>
<td>-0.047</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>X8</td>
<td>0.213</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td>X9</td>
<td>-0.265</td>
<td>0.075</td>
<td></td>
</tr>
<tr>
<td>X10</td>
<td>-0.316</td>
<td>0.062</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Comparison of predicted cost and actual cost for
validation mode

Table 8: Statistical Measures Results for Regression model

<table>
<thead>
<tr>
<th>Description</th>
<th>Statistical parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPE</td>
<td>0.824</td>
</tr>
<tr>
<td>MAPE</td>
<td>8.240</td>
</tr>
<tr>
<td>AA%</td>
<td>91.76</td>
</tr>
<tr>
<td>R</td>
<td>95.74</td>
</tr>
<tr>
<td>R²</td>
<td>93.54</td>
</tr>
</tbody>
</table>
9. Conclusions

Some conclusions resulted from achieving this research can be summarized:

1) Questionnaire is considered one of the adopted approaches in the scientific research field to better diagnose the project problems and to know the opinion of the senior management level, consultants, etc.,

2) Cost estimation traditional methods are experiencing the ill effects of the real impediments of the absence of exactness, however, a (MLR) presents a few points of interest over conventional courses for the forecast of the construction residential ventures charge estimation.

3) The conclusions from this study are:
   - Multifactor linear replace possesses the capability to evaluate the charge of building residential ventures, one an ideal determining form made as a complete apparatus for parametric cost estimation with a high level of precision 91.76%.
   - Coefficient of defining utilized to explain the linear connection amid project estimation charge and project real charge is 93.54%
   - The improved form is interesting for scholars and partners since it gives simple essay device (instrument) to forecast the charge of residential ventures.
   - The improved form of making the traditional statistics is equivalent to 20 previous residential projects as historical data and finding the connection amid output and input factors.

4) Ten impacting factors are used for prediction the total cost of residential projects, by using Multifactor Linear Regression Technique (MLRT), these factors are, the experience of the estimator, Market Status, Economic instability Accuracy of bidding documents, Site condition, Total area of the residential building, The height of residential building, Project Location, Type of residential building and Estimating method.

10. Recommendations

The researchers introducing proposals for future studies as follows:

1) Conducting similar research through employing another neural system techniques and doing comparison and evaluation among the findings to explain which technique is additional satisfactory to be utilized with the sector of management.

2) Improving many Multifactor Linear Regression models to demonstrate the methods in that many kinds of construction management difficulties ensure the winning improvement and application of this knowledge to projects problems.

3) All construction parties are encouraged to be more aware of cost estimation development and pay more attention for using a new developed technique in the estimation process.

References

