IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)
e-ISSN: 2278-1684, p-ISSN: 2320-334X
PP 60-66
www.iosrjournals.org

Mathematical Modeling to Predict Escalated Project Cost using Artificial Neural Network with an Application to Cost of Setting up Thermal Power Plant

Huzaifa A. Fidvi[1], Ravindra R.Paliwal[2], Akshyakumar S.Puttewar[3], Dr. K.N. Agrawal[4]

[1] *Assistant Professor, Mechanical Department, Anjuman College of Engineering and Technology, Nagpur.*
[2] *Principal, J.D.Polytechnic, Nagpur*
[3] *Astt. Professor, Mechanical Department, Dr.Babasaheb Ambedkar College of Engg & Research, Nagpur*
[4] *Head, Mechanical Engineering Department, Shri Ramdeobaba College of Engg & Management, Nagpur*

ABSTRACT: *The project Cost varies with time due to escalation and influence of inflationary and other conditions like demand, supply, growth, availability etc. Any business decision taken today has far reaching effects on the profitability of the project cost and thereby the viability of the organization. Various mathematical models using Time Series Methods like Linear Prediction, Trend Estimation and Causal Methods like Regression Analysis with both Linear and Non-Linear Multiple regressions were developed. This work studies the results and effectiveness of using Artificial Neural network Sigmoid Function and compares with mathematical models developed using other methods and determines the capability of ANN to indicate overruns/escalations in the project costs. Our conclusion is that, ANN gives slightly lower correlation values and higher RMS Error as it is not able to train the network with input data of only 32 data points.*

Keywords: *Artificial Neural Network, Mathematical Modelling, Pearson's correlation, Regression analysis, Trend estimation.*

I. INTRODUCTION

With time, due to inflation and other factors the value of all parameters change and the business economics valid today may not hold well tomorrow. Cost Escalation is the changes in the price of specific goods or services over a period. Various reasons of cost escalation include general money supply, changes in technology, changes in practices and imbalances in demand/supply specific to a good or service.

Construction sector largely depends upon resources such as materials, machinery and labour which are the major cost drivers of construction projects and also their prices subject to escalation. If a business proposal gets delayed due to reasons like arrangement of funds/raw material, manpower shortage, licensing and legal formalities, it results in project cost overruns.

Finding methods to both quantify and manage cost escalation on an individual project is therefore critical for owners and contractors in order to maintain profitability and ensure that there are sufficient funds to deliver the final program in budget and on schedule.

II. LITERATURE REVIEW

2.1 Introduction:

Escalation calculation by CPWD formulae

CPWD's provision for escalation allows for estimation of escalation amount using different formulae for different components of construction such as labour, material, cement and steel, P.O.L. Standard formula for all these components is as follows:

$$\mathbf{V} = \frac{W * X * (CI-CIo)}{100\ CIo}$$

Where V =variation in cost of item i.e. increase or decrease in the amount in rupees to be paid or recovered.

W = cost of work done for the period to which escalation is applicable.

X = component of item expressed as percentage of total value of work.

CI = All India Wholesale Price Index for item for period under consideration.

CIo = All India Wholesale Price Index for item as valid on the last stipulated date of receipt of tenders.

2.2 Review Of Previous Research:

Dr. N. B. Chaphalkar et al (2012) [1] studied the current structure of calculation of wholesale price index, identifies construction materials considered in the commodity basket, variation in WPI of these materials and its effect on arriving at escalation amount.

Rob J Hyndman (2009) [2] deliberates on the various methods of forecasting, steps involved in forecasting and evaluating the forecast accuracy.

Peter Morris et al (2006) [4] details methods by which participants in construction projects can both track the extent of escalation and work together to minimize the impact of cost escalation on the success of a project.

Touran Ali, Ramon Lopez (2006) [5] studied the methods of modeling cost escalation in large infrastructure projects.

Williams T. (1994) [6] examines predicting changes in construction cost indexes using neural networks.

C.M. Bishop, C.M. Roach. (1992) [7] examined the application of Neural networks and concludes that it is a tool for the fast solution of repetitive nonlinear curve fitting problems.

Huzaifa A. Fidvi et al [8] developed various mathematical models for estimating the escalated project using trend line, simultaneous equation and factor based regression analysis.

III. OBJECTIVE

The objective of this paper is to compare different mathematical model obtained by various trend line, regression methods with the results obtained by Artificial Neural Network to ascertain the effectiveness of using ANN for Prediction of Escalated Project Cost of Setting up Thermal Power Plant.

IV. METHODOLOGY

Various parameters are selected that have a bearing on industry and economy. Economic Parameters like inflation Rate, Consumer Price Index - (CPI), Wholesale Price Index - (WPI) and Industrial Parameters like Energy Prices (Electricity), Steel Rate (MS 1020), Labour Rate (Min Wages), Cement Rates, Coal Mining, Cost of Capital Goods - Machinery.

The various Quantitative methods of forecasting include the Time Series Methods like Autoregressive Moving Average (ARMA), Exponential Smoothing, Extrapolation, Linear Prediction, Trend Estimation, Growth Curve and Explanatory Models which include Causal Methods, Linear and Non-Linear Regression Analysis and Econometrics. The methods are appplied depending upon the applicability of each model to the parameters.

Artificial Neural Network can be used to get more reliable forecast values and trend estimation. The neural network algorithm is typically much faster than interactive approaches.

Applying Neural Network:

Introduction - Neural networks are the preferred tool for many predictive data mining applications because of their power, flexibility, and ease of use.

A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

• Knowledge is acquired by the network through a learning process.

• Interneuron connection strengths known as synaptic weights are used to store the knowledge.

The traditional linear regression model can acquire knowledge through the least-squares method and store that knowledge in the regression coefficients. In this sense, it is a neural network. In fact, we can argue that linear regression is a special case of certain neural networks. However, linear regression has a rigid model structure and set of assumptions that are imposed before learning from the data.

A neural network can approximate a wide range of statistical models without requiring that we hypothesize in advance certain relationships between the dependent and independent variables. Instead, the form of the relationships is determined during the learning process. If a linear relationship between the dependent and independent variables is appropriate, the results of the neural network should closely approximate those of the linear regression model. If a nonlinear relationship is more appropriate, the neural network will automatically approximate the "correct" model structure.

The trade-off for this flexibility is that the synaptic weights of a neural network are not easily interpretable. Thus, if we are trying to explain an underlying process that produces the relationships between the dependent and independent variables, it would be better to use a more traditional statistical model. However, if model interpretability is not important, we can often obtain good model results more quickly using a neural network.

V. APPROACH AND ANALYSIS

Various parameters, methods used and mathematical models developed as referred to Huzaifa Fidvi et al [8] *"Development of Mathematical Model to Predict the Project Cost in Case of Time Overruns"* are studied and summarized below.

5.1 **Mathematical Modeling of Industrial Parameters:**

Curve fitting was done using the linear and polynomial regression techniques on MS EXCEL software.

Analysis was carried out to develop various mathematical models [8] using Linear and Polynomial Regression and trend lines plotted using the actual data from 1981 to 2010 and the following mathematical models developed establishing relationships between dependent variable 'price of the industrial parameter' and independent variable 'time' of the various industrial parameters. The table shows their respective values of 'coefficient of Determination R^2'.

Table 2: Summary of Mathematical Models of Industrial Parameters

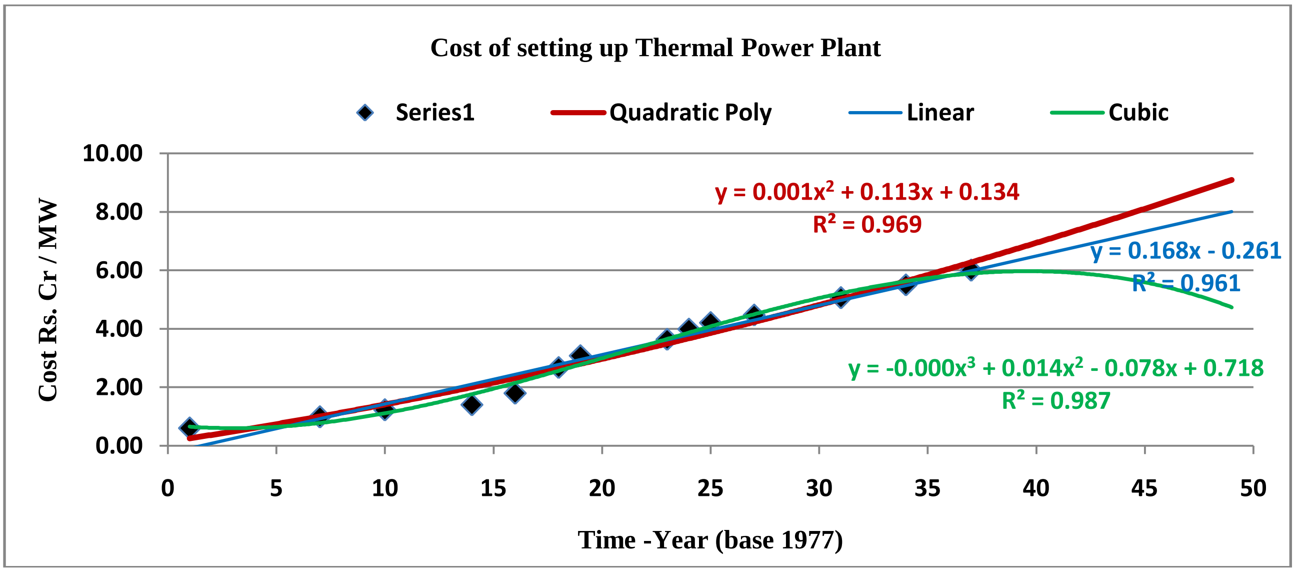
S. No	Parameter	Mathematical Model	Output Equation	R^2
1	Cement WPI	Cubic Polynomial Regression	$Yc= 0.0135year^3 - 0.3175year^2 + 9.55year + 97.749$	0.9527
2	Coal WPI	Quad Polynomial Regression	$Yc= 0.6646year^2 + 9.5037year + 82.522$	0.9939
3	Electricity WPI	Quad Polynomial Regression	$Yc = 0.3628year^2 +31.863Year - 9.8657$	0.9723
4	Machinery WPI	Quad Polynomial Regression	$Yc = 0.0518Year^2 - 11.326Year + 65.964$	0.9827
5	Steel WPI	Cubic Polynomial Regression	$Yc = 0.047year^3 - 1.3593Year^2 - 21.276Year + 54.696$	0.9631
6	Wages Index	Cubic Polynomial Regression	$Yc = 0.3097year^3 - 8.008year^2 + 93.8834year - 94.0112$	0.9783
Base Year - 1981 where (year=1, 2, 3….. for 1981, 1982, 1983…..)				

Using these mathematical models forecasting of the price of the industrial parameters WPI is done for a further period of 15 years i.e. up to 2025.

5.2 Mathematical Models [8] to Predict the Project Cost In Case of Time Overruns.

To develop the mathematical model for estimating the effect of Time overruns on the Project Cost the following approaches were taken.

5.2.1 Trend line Based Polynomial Regression Approach:-



5.2.2 Factor based Regression: -

5.2.2.1 Simultaneous Equations Approach.

For Setting up a Thermal Power Plant the following factors are considered: Cement, Coal, Electricity for Industry, Machinery and Machine Tools, Steel Rates, Wages.

COST = f(X_1, X_2, X_3, X_4, X_5, X_6)

$$COST=A_1X_1+A_2X_2+A_3X_3+A_4X_4+A_5X_5+A_6X_6$$

Where X_1=Cement, X_2=Coal, X_3=Electricity, X_4=Machinery, X_5=Steel, X_6=Wages and A_1, A_2, A_3, A_4, A_5, A_6 are coefficients

The six Simultaneous Equations formed using values for different years are solved using Matrices and values of the coefficients A_1, A_2, A_3, A_4, A_5 and A_6 are obtained.

5.2.2.2 Regression Analysis Approach.

Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships.

The general form for a multiple linear regression equation is:

$$Yc = a + b_1X_1 + b_2X_2 + ... + b_kX_k$$

Where Yc = calculated (predicted) value of the dependent variable, a = intercept (constant term)

X_j = jth independent (predictor) variable, b_j= coefficient associated with the jth indpt variable.

5.3 Summary of Mathematical Models for calculating Cost of Setting up Thermal Power Plant.

5.3.1 Trend line based - Quadratic Polynomial Regression Approach

Cost Yc = $0.0014x^2 + 0.1137x + 0.134$

Where x = 1, 2, 3,.........for years 1977, 1978, 1979,........

5.3.2 Factor Based Regression - Simultaneous Equations Approach.

COST Yc = $-0.0351*X_1 – 0.008369*X_2 – 0.000425*X_3 +0.04288*X_4 – 0.00723*X_5 + 0.001091*X_6$

Where: X_1-Cement = $0.0135*year^3 - 0.3175*year^2 + 9.55*year + 97.749$

X_2-Coal = $0.6646*year^2 + 9.5037*year + 82.522$

X_3-Electricity = $0.3628*year^2 + 31.863*Year - 9.8657$

X_4-Machinery = $0.0518*Year^2 - 11.326*Year + 65.964$

X_5-Steel = $0.047*year^3 - 1.3593*Year^2 - 21.276*Year + 54.696$

X_6-Wages = $0.3097*year^3 - 8.008*year^2 + 93.8834*year - 94.0112$

Base Year - 1981 (year=1, 2, 3….. for 1981, 1982, 1983…..)

5.3.3 Factor Based Approach - Using Regression Analysis.

COST Yc = -0.00096*X1 + 0.00033*X2 + 0.00144*X3 + 0.00864*X4 - 0.00035*X5 + 0.000048*X6 + 0.10821

VI ANN BASED APPROACH

6.1 Mathematical Model : ANN Based Approach Using Fitnet Network

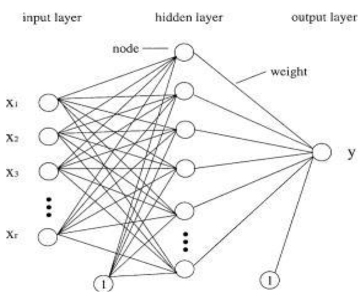
The input data for ANN is same as that used for Regression Analysis. The entire analysis of ANN is done on MATLAB R2011b version. The architecture used and the Matlab training output is shown below.

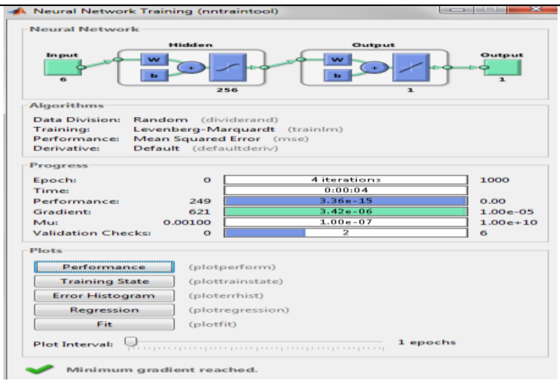
6.1.1 Architecture of Neural Network:

Software	-	MATLAB R2011b version
Network	-	'Fitnet' is used for function Mapping & Fitting.
Data	-	Normalized
Input Layer	-	One having 6 neurons with 32 training inputs
Hidden Layer	-	One having 256 neurons
Output Layer	-	One having SIGMOID FUNCTION (tansig)
Output	-	Single, Cost factor

Image 1 ANN Architecture **Image 2 Neural Network Training Output - MATLAB**

IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)
e-ISSN: 2278-1684, p-ISSN: 2320-334X
PP 60-66
www.iosrjournals.org





6.2 Network Training and Output.

After training the network the following outputs are derived.

Table 3: Neural Network Output, Residuals and %Error of Cost

NEURAL NETWORK – COMPUTED COST Rs. Cr per MW									
Year	**Actual Cost Ya**	**Computed Cost Yc**	**Residuals**	**%Error**	**Year**	**Actual Cost Ya**	**Computed Cost Yc**	**Residuals**	**%Error**
1981	0.74	**1.14**	-0.40	**-54.4**	***1997***	3.32	**3.02**	0.30	**9.1**
1982	0.87	**1.41**	-0.54	**-62.8**	***1998***	3.64	**2.91**	0.73	**19.9**
1983	0.98	**1.43**	-0.45	**-45.6**	***1999***	3.86	**3.13**	0.73	**18.8**
1984	1.13	**1.69**	-0.55	**-48.7**	***2000***	3.99	**2.82**	1.17	**29.4**
1985	1.27	**1.92**	-0.65	**-50.9**	***2001***	4.2	**3.18**	1.02	**24.2**
1986	1.22	**1.60**	-0.38	**-31.4**	***2002***	4.04	**3.47**	0.58	**14.3**
1987	1.56	**1.59**	-0.04	**-2.3**	***2003***	4.47	**3.39**	1.08	**24.2**
1988	1.7	**1.61**	0.09	**5.4**	***2004***	4.42	**3.43**	0.99	**22.5**
1989	1.85	**1.47**	0.38	**20.8**	***2005***	4.62	**3.56**	1.06	**22.9**
1990	2	**1.69**	0.31	**15.5**	***2006***	4.82	**3.89**	0.93	**19.3**
1991	2.16	**2.07**	0.09	**4.1**	***2007***	5.06	**4.60**	0.46	**9.1**
1992	2.31	**2.10**	0.22	**9.4**	***2008***	5.22	**5.21**	0.01	**0.2**
1993	2.47	**2.29**	0.19	**7.6**	***2009***	5.42	**5.21**	0.22	**4.0**
1994	2.68	**2.30**	0.37	**13.9**	***2010***	5.5	**5.21**	0.29	**5.3**
1995	3.08	**2.69**	0.39	**12.6**	***2011***	5.63	**5.42**	0.21	**3.8**
1996	3.14	**3.23**	-0.09	**-2.9**	***2012***	5.84	**5.84**	0.00	**0.0**

6.3 Descriptive Measures of Forecast Accuracy Developed using ANN:

MAD = 0.26, MSE =0.32, RMSE = 0.565, Pearson's Correlation Factor between Ya and Yc = 0.9534

6.4 Validation and Prediction of Cost: using ANN Fitnet Network

Table 4 Validation of Project Cost – ANN based Approach using Fitnet Network

VALIDATION									
Year	X1 Cement	X2 Coal	X3 Electricity	X4 Machinery	X5 Steel	X6 Wages	Actual Cost Ya	Computed Cost Yc	% Error
2013	553.1	1119.8	1436.7	496.1	966.2	5413.3	**6.00**	**6.30**	**- 5.00**

Table 5 Prediction of Project Cost – ANN based Approach using Fitnet Network

FORECASTS							
Year	**X1 Cement**	**X2 Coal**	**X3 Electricity**	**X4 Machinery**	**X5 Steel**	**X6 Wages**	**Computed Cost Yc**
2015	622.8	1229.2	1549.8	525.8	1150.1	6660.7	**7.09**
2020	837.1	1526.0	1845.1	601.9	1740.0	10669.6	**9.54**
2025	1116.7	1855.9	2158.6	680.5	2544.0	16136.4	**12.72**

VII. RESULTS

Summary of results obtained from different approaches used for Mathematical Modeling are detailed as under.
Table 6 Validation and Prediction of Cost for setting up a Thermal Power Plant

VALIDATION											
Year	**X1 Cement**	**X2 Coal**	**X3 Electric**	**X4 M/C**	**X5 Steel**	**X6 Wages**	**COST Rs. Cr per MW**				
							Ya	**Trend Line Model**	**Factor Based Model**	**Regression Model**	**ANN**
2013	553.1	1119.8	1436.7	496.1	966.2	5413.3	**6.0**	**6.06**	**6.00**	**6.23**	**6.30**
PREDICTION											
2015	622.8	1229.2	1549.8	525.8	1150.1	6660.7	**-**	**6.72**	**6.65**	**6.61**	**7.09**
2020	837.1	1526.0	1845.1	601.9	1740.0	10669.6	**-**	**7.87**	**8.68**	**7.58**	**9.54**
2025	1116.7	1855.9	2158.6	680.5	2544.0	16136.4	**-**	**9.10**	**11.39**	**8.53**	**12.72**

The Accuracy of Mathematical Models are determined as follows:
Table 7 Summary of Descriptive Measures of Accuracy of Cost

Descriptive Measures	**MATHEMATICAL MODEL**			
	ANN	**TREND LINE**	**FACTORS Based**	**REGRESSION**
MD	**0.26**	**0**	**0.64**	**0.01**
MAD	**0.46**	**0.25**	**0.71**	**0.10**
MSE	**0.32**	**0.0897**	**0.9162**	**0.0139**
RMSE	**0.565**	**0.2996**	**0.9572**	**0.118**
Pearson's Correlation (r)	**0.9534**	**0.9846**	**0.9281**	**0.9972**

VIII. CONCLUSION

The variation of Economic and Industrial parameters under inflationary conditions and also due to number of factors at play means that virtually no currently published index provides a reliable method of tracking or predicting future trends.

The behaviour of all selected Industrial parameters was studied and their variation over time, their interdependence, and best fit curve analysed. Mathematical Model for industrial parameters were formed and future values forecasted.

Relevant parameters are used to formulate mathematical model to indicate cost overruns/escalations in the project cost for setting up a thermal power plant in case of time overruns. Among all the approaches taken the factor based Regression Mathematical Model is giving the best results with very low (a) Mean Deviation of 0.01 (b) Mean Absolute Deviation of 0.10 (c) Root Mean Square Error of 0.118 Rs. Crore per MW and a very high 0.9972 Pearson's Correlation Coefficient between the Actual and Computed values.

ANN was developed using Fitnet software in MATLAB. Correlation between Ya & Yc is 0.9534 (comparatively lower) and Root Mean Square Error is Rs. 0.565 crore per MW (quite high) were obtained. the software is not able to train the network with input data of only 32 points, hence ANN is not effective. In the given circumstances the factor based Regression Mathematical Model is accepted as it is giving best results.

REFERENCES

Dr. N. B. Chaphalkar, Ms. Sayali S. Sandbhor (2012) "Wholesale Price Index and its effect on Price Escalation of materials for Indian construction industry" (IJERT) Vol. 1 Issue 3, May – 2012.
Rob J Hyndman (2009) "Forecasting overview".
Paras Mandal et al (2007) - "A Novel Approach to Forecast Electricity Price for PJM using Neural Network and Similar Days Method." 2058 IEEE Transactions on Power Systems, Vol. 22, No. 4, November 2007
Peter Morris, William F. Willson (2006) "Measuring and Managing Cost Escalation" 2006 AACE International Transactions CSC.06
Touran Ali, Ramon Lopez (2006), "Modeling cost escalation in large infrastructure projects", Journal of construction engineering and management, ASCE, Vol. 132, No. 8, pp. 853-860.
Williams T. (1994), 'Predicting changes in construction cost indexes using neural networks', Journal of construction engineering and management, ASCE, Vol. 120, No. 2, pp. 306-320.
C.M. Bishop, C.M. Roach. (1992). "Fast curve fitting using neural networks".
Huzaifa Fidvi et al (2013) "Development of Mathematical Model to Predict the Project Cost in Case of Time Overruns".
Neal Wagner et al (2006) - "Time Series Forecasting for Dynamic Environments: The DyFor Genetic Program Model."
Xiaohong Chen (2001) - "Semiparametric ARX Neural-Network Models with an Application to Forecasting Inflation"

John H Cochrane-"Time Series for Macro Economics and Finance".
Oxford University Press – Economic Survey 2011-2012.
Data Source: -- - Central Statistics Office (CSO).
i.)Directorate of Economics & Statistics. ii). Reserve Bank of India. iii).Economic Adviser, Ministry of Commerce and Industry. iv) Directorate of Economics & Statistics. v).Ministry of Finance (Department of Economic Affairs). vi). Department of Public Enterprises & Ministry of Labour.
The Senior Manager(F & A), Thermal Power Station, Khaparkheda.
Data of NTPC Power Plants (All India).