

Scientific Review – Engineering and Environmental Sciences (2020), 29 (2), 234–243  
Sci. Rev. Eng. Env. Sci. (2020), 29 (2)  
Przegląd Naukowy – Inżynieria i Kształtowanie Środowiska (2020), 29 (2), 234–243  
Prz. Nauk. Inż. Kszt. Środ. (2020), 29 (2)  
<http://iks.pn.sggw.pl>  
DOI 10.22630/PNIKS.2020.29.2.20

**Firas Kh. JABER<sup>1</sup>, Nidal A. JASIM<sup>2</sup>, Faiq M.S. Al-ZWAINY<sup>3</sup>**

<sup>1</sup>Middle Technical University, Electrical Engineering Technical College

<sup>2</sup>University of Diyala, College of Engineering

<sup>3</sup>Al-Nahrain University, College of Engineering

## **Forecasting techniques in construction industry: earned value indicators and performance models**

**Key words:** Machine Learning Regression Techniques (MLRT), earned value indexes, SPI, CPI, and TCPI

### **Introduction**

The construction industry is an important industry for any government due to its direct association with the implementation of the goals and policies of the government in various fields of concern to the citizen in terms of education, health, housing and other facilities and services. The construction industry is also one of the broad and important sectors of any country's economy, and it is one of the main engines that governments resort to move the economy and create jobs and reduce unemployment. And the construction industry has unique characteristics, which made it an industry that is distinguished from the rest of the

other industries by its many risks, and its projects always suffer from the problems of delay in implementation and increase in cost in most countries of the world. Among the most important characteristics of the construction industry (Myers, 2005):

1. The nature of its product is unique, each project differs from the other, and the temporary of each project is limited in duration and location, with the completion of the project, the equipment and labor will be transferred to another project in another place.
2. The nature of work within a single project is fragmented, as several different parties separate and separate to complete the project.
3. The industry's heavy dependence on manpower and its lack of research and development in methods and methods of implementation.

4. The criterion of competition for bidding in most cases is the lower price, which reduces the profit margin and thus reduces the desire of the companies executing in research, development and creativity.

The increasing demand for tall buildings represents a huge opportunity to reconstruct the construction industry and create value through joint action. According to research prepared by the McKinsey Global Institute, the industry can boost the productivity of its workers by up to 60% if changes are made in seven key areas: regulations, design operations, contracts, supply and supply chain management, and on-site implementation (Granskog, Guttman & Sjödin, 2016).

Tall buildings since the 1990s and during the first decade of the 21st century are considered civilization and urbanization facilities, not only that, but tall buildings have become symbols of countries. Tall buildings are defined as those buildings whose height is not less than 40 m or 13 floors. In most cases, the management of tall projects is not a simple matter, as success in such projects requires good coordination as well as a comprehensive integrated operating system for the project, so that all project participants can understand their roles and agree on key performance indicators (Al-Zwainy & Edan, 2017).

Earned value management system (EVMS) is a useful management technique available for tall buildings projects managers to monitor and control projects; a technique that combines the work scope, schedule and the cost elements of a project and facilitates the integrated reporting of a project's progress and the

cost status. Earned value management system in effect integrates time management and cost management which are essential elements of tall buildings projects management (Al-Zwainy, Mohammed & Mohsen, 2015).

Review of the literature related to earned value management system with Forecasting techniques such as Machine Learning Regression Techniques (MLRT) was in top management journals such as International Journal of Project Management. Where, various researchers have used Machine Learning Regression Techniques (MLRT) as a tool for prediction and optimization in different project management areas are also reviewed, including earned value management. The great majority of project management applications of Multiple Linear Regression are based on the inter algorithm. The previous studies are among the most important scientific foundations on which the current study is based. The researcher begins with research, examination and analysis in the previous studies; this is because the research processes are cumulative processes that depend on the previous ones. In the next paragraph, the previous studies on the earned value management will be addressed and techniques used for forecasting purposes.

During the last few years or so, the use of EVM has increased in many construction engineering projects and has demonstrated some degree of success. The following literature review reveals that EVM have been used successfully in modelling of the earned value management. Pajares and Lopez-Paredes (2011) proposed two new metrics that combine earned value management and project

risk management for construction project controlling and monitoring. Elshaer (2013) studied the impact of the activities sensitivity information on the forecasting accuracy of the Earned Schedule Method (ESM). Czemplik (2014) applied schedule forecast Indicator (SFI) to support site managerial decision concerning variation orders. Khamooshi and Golafshani (2014) provided a new approach to decouple the schedule and cost dimensions in EVM by adding earned duration while maintaining the unique interaction of the three major project management elements of scope, cost, and time similar to EVM. Chen, Chen and Lin (2016) improved the predictive power of Planned Value (PV) before executing construction project. Jaber, Hachem and Al-Zwainy (2019) applied the value management methodology acquired in the waste water treatment plant projects to measure performance by predicting performance indicators such as CPI, SPI and TCPI.

The following literature review reveals that Machine Learning Regression Techniques (MLRT) have been used successfully in construction project management (Al-Zwainy et al., 2015) developing EAC model for bridges projects using Multifactor Linear Regression Technique (MLRT). Bilal and Oyedele (2020) presented guidelines for Applied Machine Learning (AML) in the construction industry from training to operationalizing models, which are drawn from our experience of working with construction folks to deliver Construction Simulation Tool (CST). Sabahi and Parast (2020) used predictive analytics by proposing a machine learning approach to predict individuals' project performance based on measures of several aspects of entrepre-

neurial orientation and entrepreneurial attitude of the individuals.

The researchers did not find any study that is exactly the same as the current study while searching in international libraries and discreet periodicals such as Scopus, Springer, Taylor France and others, as this current study is almost unique of its kind that deals with predicting the value indicators gained using Machine Learning Regression Techniques (MLRT) in tall building projects, despite the existence of multiple studies dealing with a topic prediction in construction projects. Based on the above discussion the contribution of this study are as follows.

Firstly, in this study, the main target is on using associate Forecasting techniques for the aim of earned value prediction, meant as a decision-support aid to project manager, contractors and planners. The contribution of this study is to spot the most effective activity in terms of model structure and parameters on an earned value prediction. Applied interest of this study may well be use machine intelligence techniques such as Machine Learning Regression Techniques (MLRT) as an acceptable tool for earned value estimating, where, Project managers in Iraq; typically have to be compelled to estimate the earned value of construction projects at construction stage quickly and around to supply funding or to get the adoption of the budget from decision-makers. Therefore, it's necessary to search out the simplest way to understand the earned value of the construction projects in a very short time with acceptable accuracy.

In summary, the application EVM in tall building projects would be specifi-

cally helpful for a state like Iraq which suffers from shortage of a high number of severe tall buildings. The main objective of this study is to develop Machine Learning Regression Techniques (MLRT) to predict the earned value indicators at early stage of tall buildings projects in Iraq to reduce the percentage error of estimation. This will facilitate the project practitioners to the further use of EVA for the schedule and cost control of their construction projects. To achieve this, there is a need to identify the factors that affect the performance of residential buildings projects that can be available at early stages. Therefore, the specialists researchers in this study are endeavoring improvement and assessment earned value models through the below stages:

1. Choose the appropriate statistics software.
2. Identification of MLRT variables that have an effect on the earned value index in tall buildings projects.
3. Building a numerical model to foresee of the earned value indexes (SPI, SPI and TCPI).
4. Check the confirmation and approval of the scientific models.

### **Choose the appropriate statistics software**

Several applications that support the establishment of statistics analysis like Microsoft Excel, Statistica, MINITAB, and MATLAB, but this study was selected SPSS Program, where SPSS is the premier statistics analysis environment. The SPSS range provides an easy-to-use, visual, object-oriented approach to problem solving using intelligent technologies.

SPSS is short for Statistical Package for the Social Sciences, and it's utilized by different sorts of analysts for complex measurable information investigation. The SPSS programming bundle was made for the administration and factual examination information in different sciences. Long produced by SPSS Inc., it was acquired by IBM in 2009. The current versions are named IBM SPSS Statistics.

### **Identification ANN Models variables**

The Bismayah City Project is the first of its kind and the largest in the history of Iraq in terms of size and level of services and infrastructure provided in it, and Hanwha Engineering and Construction Company is proud to be the first company to implement such a strategic project in Iraq. The new city of Basmaya is an integrated residential complex located 10 km southeast of the capital, Baghdad, Iraq, on the international road No 6 linking the capital Baghdad and Wasit Governorate, and on an area of 1,830 ha, and consists of 100,000 housing units, and also includes a network Integrated infrastructure of water and electricity, as well as schools and recreational and commercial complexes. The cost of constructing the complex amounts to \$7.75 billion, the project is implemented by the Korean Hanwha Engineering and Construction Company, which works in the field of construction and construction around the world, with a duration of implementation of up to seven years. The new Basmaya city represents a new modern lifestyle that may be characterized by

a high level of services and an ideal environment suitable for the development of new generations. The aerial photos show the level of progress of work and completion of other residential and service buildings, as shown in Figure 1.



FIGURE 1. The new city of Basmaya in Iraq

Machine Learning Regression Techniques (MLRT) requires a lot of data. Therefore, a lot of historical tall buildings projects were collected which were done between 2012 and 2018 as shown in Table 1. There are two types of variables that affected on the earned value in tall buildings projects in Republic of Iraq which are dependent variables and independent variables. Cost Performance Index

TABLE 1. Summary of historical data for Bismayah City Project

Parameter	Max.	Min.	Range	AVG
BAC	950	800	150	865
AC	600	200	400	375
A%	60	25	35	39
EV	510	237.5	272.5	339.8
P%	60	25	35	42
PV	510	237.5	272.5	358
SPI	1.11	0.75	0.36	0.95
CPI	1.18	0.80	0.38	0.93
TCPI	0.80	1.19	0.80	0.39

index (CPI), Schedule Performance Index (SPI) and To Complete Cost Performance Indicator (TCPI) are defined as the dependent variables. There are many variables as an independent variable such as: Budget at Completion (BAC), Actual Cost (AC), Actual Percentage (A%), Earned Value (EV), Planning Percentage (P%), and Planning Value (PV).

### Building Machine Learning Regression (MLR) Models

The factors that were recognized at the information distinguishing proof stage were utilized to build up the Machine Learning Regression Techniques (MLRT). Three numerical models were created in this examination by utilizing Machine Learning Regression Techniques (MLRT). It was utilized on the grounds that it is probably the most kind that broadly utilized in this field of study. What's more, the venture qualities or parameters in a numerical model were utilized to foresee earned value indexes. The Statistical Product and Solutions Services (SPSS) programming forms 26 are utilized as a tool to build up the three models, as following:

1. Schedule Performance Index (SPI).
2. Cost Performance Index (CPI).
3. To Complete Cost Performance Indicator (TCPI).

### First: Schedule Performance Index Model

Table 2 expounds multiple linear regression estimates, such as coefficients of standardized, coefficients of unstandardized, standardized errors, t-test and

TABLE 2. Coefficients of SPI-Model (dependent variable:  $Y_1$ )

Model	Unstandardized coefficients		Standardized coefficients	$t$	sig.
	$B$	$SE$	$\beta$		
Constant	-14.393	8.593	-	-1.675	0.193
$X_1$	0.018	0.010	3.253	1.784	0.172
$X_2$	0.007	0.005	2.684	1.255	0.298
$X_3$	-0.076	52.969	-0.026	-0.001	0.999
$X_4$	-0.013	0.066	-3.922	-0.197	0.857
$X_5$	33.633	50.328	10.067	0.668	0.552
$X_6$	-0.034	0.057	-8.747	-0.590	0.597

testing the significant of every autonomous factors or independent variable.

The regression statistics of Schedule Performance Index Model (SPI-Model) are set in Table 1, which can be formed as Equation (1):

$$Y_1 = SPI = -14.393 + (0.018X_1) + (0.007X_2) - (0.076X_3) - (0.013X_4) + (33.633X_5) - (0.034X_6) \quad (1)$$

### Second: Cost Performance Index Model

The regression statistics of Cost Performance Index Model (CPI-Model) are set in Table 3.

Statistics of regression for model CPI are set in Table 3, which can be formed as Equation (2):

$$Y_2 = CPI = -12.405 + (0.015X_1) + (0.002X_2) - (22.344X_3) + (0.016X_4) + (50.059X_5) - (0.05X_6) \quad (2)$$

### Third: Complete Cost Performance Indicator

The regression statistics of To Complete Cost Performance Indicator (TCPI-Model) are set in Table 4.

TABLE 3. Coefficients of CPI-Model (dependent variable:  $Y_2$ )

Model	Unstandardized coefficients		Standardized coefficients	$t$	sig.
	$B$	$SE$	$\beta$		
Constant	-12.405	7.813	-	-1.588	0.211
$X_1$	0.015	0.009	2.609	1.641	0.199
$X_2$	0.002	0.005	0.881	0.472	0.669
$X_3$	-22.344	48.16	-7.317	-0.464	0.674
$X_4$	0.016	0.060	4.731	0.272	0.803
$X_5$	50.059	45.75	14.372	1.094	0.354
$X_6$	-0.050	0.052	-12.444	-0.963	0.407

TABLE 4. Coefficients of CPI-Model (dependent variable:  $Y_3$ )

Model	Unstandardized coefficients		Standardized coefficients	$t$	sig.
	$B$	$SE$	$\beta$		
Constant	-1.436	8.477	-	-0.169	0.876
$X_1$	0.001	0.010	0.164	0.092	0.932
$X_2$	-0.002	0.005	-0.789	-0.378	0.731
$X_3$	-31.521	52.258	-10.647	-0.603	0.589
$X_4$	0.036	0.065	10.635	0.547	0.623
$X_5$	31.441	49.653	9.311	0.633	0.572
$X_6$	-0.030	0.057	-7.538	-0.521	0.638

Statistics of regression for model TCPI are set in Table 4, which can be formed as Equation (3):

$$Y_3 = TCPI = -1.436 + (0.001X_1) - (0.002X_2) - (31.521X_3) + (0.036X_4) + (31.441X_5) - (0.03X_6) \quad (3)$$

### Check the Confirmation and Approval of the Scientific Models

Below is the model's summary, which has some important statistics. It provides us correlation coefficient ( $R$ ) with determination coefficient ( $R^2$ ) and

the estimation of standard error. This statistical analysis was bring about for the MLR models (SPI, CPI, and TCPI) between output (earned value indexes) and input (BAC, Budget at Completion -  $X_1$ ; AC, Actual Cost -  $X_2$ ; A%, Actual Percentage -  $X_3$ ; EV, Earned Value -  $X_4$ ; P%, Planning Percentage -  $X_5$ ; PV, Planning Value -  $X_6$ ). On the other hand, the  $R$ -values for SPI, CPI, and TCPI models are equal to 85.5%, 89.2%, and 86.3% respectively, which is viewed as extremely aloft correlation. In addition, the determination coefficient shows the extent of the variety in input variable from an output variables, and that models have

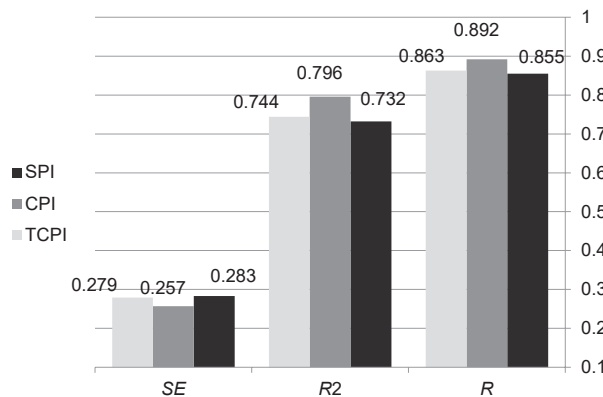


FIGURE 1. Summary statistics of models

$R^2$  values equal to 73.2%, 79.6%, and 74.4% respectively.

From Figure 1, it can be seen that model (SPI), model (CPI), model (TCPI) verification have pretty performance in order that offered high correlation of coefficient ( $R$ ) 85.5%, 89.2% and 86.3%, respectively. Figures 2–4 illustrate the capability of MLR models (SPI-model, CPI-model, TCPI-model), where, the coefficient of determination ( $R^2$ ) is 73.2%,

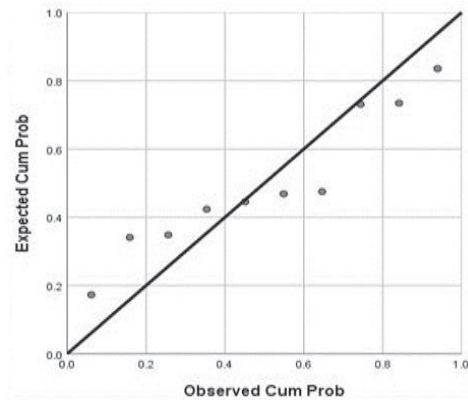


FIGURE 2. Relationship between observed SPI and expected SPI

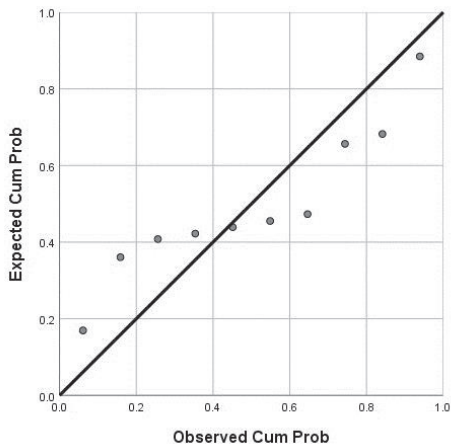


FIGURE 3. Relationship between observed CPI and expected CPI

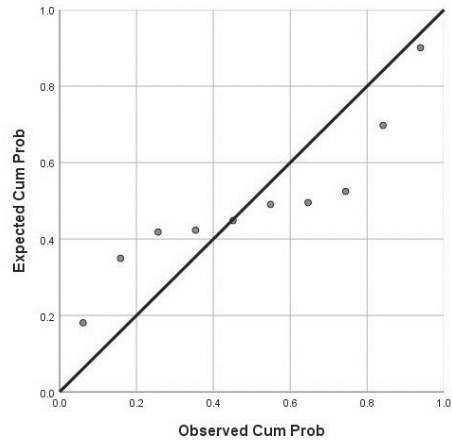


FIGURE 4. Relationship between observed TCPI and expected TCPI

79.6% and 74.4%, therefore, it can be concluded that these models shows an excellent agreement with the actual measurements.

## Conclusions

In this study, a depiction of models factors, information assortment strategy and factual determination of verifiable information has been clarified. Bismayah New City was selected as a case study, six variables were adopts as independent variables (BAC, Budget at Completion –  $X_1$ ; AC, Actual Cost –  $X_2$ ; A%, Actual Percentage –  $X_3$ ; EV, Earned Value –  $X_4$ ; P%, Planning Percentage –  $X_5$ ; PV, Planning Value –  $X_6$ ), and three variables were (Cost Performance Index – CPI; Schedule Performance Index – SPI; To Complete Cost Performance Indicator – TCPI) are defined as the dependent variables. In addition, Machine Learning Regression Techniques (MLRT) has been used to build prediction model for earned value indexes by



using SPSS-26 statistically software. It was found that the MLRT showed good results of estimation in terms of correlation coefficient ( $R$ ) generated by MLR models for SPI and CPI and TCPI where the  $R$  were 85.5%, 89.2%, and 86.3% respectively. At long last, a result tends to be presumed that these models show a brilliant concurrence with the genuine estimations.

### The availability of data

The study discoveries which were bolstered by the information, are realistic by the relating's essayist can be obtained on demands.

### Conflicts of interest

The authors asserted that they hadn't interest clashes and conflicts.

### Acknowledgements

Authors would like to thank those respected persons who have helped and guided us to achieve this paper; they really deserve all respect and gratitude. Then, we'd like to thank all the staff of the Al-Nahrain University, Middle Technical University, Baghdad, and the University of Diyala in Iraq to help us to get guideline for our paper by providing us lots of consultations. As well as, we also want to thank and give our gratitude to everyone who has guided or advised us in submitting or writing this paper.

### Author contributions

Firas Kh. Jaber conceived of the presented idea and contributed to sample preparation and wrote the manuscript. Faiq M.S. Al-Zwainy developed the MLR models and performed the com-

putations. Nidal A. Jasim confirmed the logical techniques and regulated the discoveries of this work. All researchers talked about the outcomes and writing to the last composition of manuscript.

### References

- Al-Zwainy, F.M.S. & Edan, I.A. (2017). *Information technology in construction project management*. Amman, Jordan: Darghaidaa for Publishing [translated from Arabic].
- Al-Zwainy, F.M.S., Mohammed, I. & Mohsen, D.S. (2015). *Earned value management in construction project*. Saarbrücken: LAP LAMBERT Academic Publishing.
- Bilal, M. & Oyedele, L.O. (2020). Guidelines for applied machine learning in construction industry – a case of profit margins estimation. *Advanced Engineering Informatics*, 43, 101013. DOI 10.1016/j.aei.2019.101013
- Chen, H.L., Chen, W.T. & Lin, Y.L. (2016). Earned value project management: improving the predictive power of planned value. *International Journal of Project Management*, 34(1), 22-29.
- Czemplik, A. (2014). Application of earned value method to progress control of construction projects, *Procedia Engineering*, 91(1), 424-428.
- Elshaer, R. (2013). Impact of sensitivity information on the prediction of project's duration using earned schedule method. *International Journal of Project Management*, 31(4), 579-588.
- Granskog, A., Guttman, B. & Sjödin, E. (2016). *Toward a customer-centric construction-equipment industry*. McKinsey & Company. Retrieved from: <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/toward-a-customer-centric-construction-equipment-industry>
- Jaber, F.K., Hachem, S.W. & Al-Zwainy, F.M. (2019). Calculating the indexes of earned value for assessment the performance of waste water treatment plant. *ARP Journal of Engineering and Applied Sciences*, 14(4), 792-802.

- Khamooshi, H. & Golafshani, H. (2014), EDM: Earned Duration Management, a new approach to schedule performance management and measurement. *International Journal of Project Management*, 32(6), 1019-1041.
- Myers, D. (2005). *Construction economics: a new approach*. London: Spon Press.
- Pajares, J. & Lopez-Paredes, A. (2011). An extension of the EVM analysis for project monitoring: The Cost Control Index and the Schedule Control Index. *International Journal of Project Management*, 29(5), 615-621.
- Sabahi, S. & Parast, M.M. (2020). The impact of entrepreneurship orientation on project performance: a machine learning approach. *International Journal of Production Economics*, 107621. DOI 10.1016/j.ijpe.2020.107621

see Schedule Performance Index (SPI) as first model, Cost Performance Index (CPI) as a second model, and the third model is To Complete Cost Performance Indicator (TCPI) in Bismayah New City was chosen as a case study. The methodology is mainly impacted by the deciding various components (variables) which impact on the earned value analysis, six free factors ( $X1$ : BAC, Budget at Completion;  $X2$ : AC, Actual Cost;  $X3$ : A%, Actual Percentage;  $X4$ : EV, Earned Value;  $X5$ : P%, Planning Percentage, and  $X6$ : PV, Planning Value) were self-assertively assigned and agreeably depicted for per tall buildings projects. It was found that the MLRT showed good results of estimation in terms of correlation coefficient ( $R$ ) generated by MLR models for SPI and CPI and TCPI where the  $R$  were 85.5%, 89.2%, and 86.3% respectively. At long last, a result tends to be presumed that these models show a brilliant concurrence with the genuine estimations.

## Summary

**Forecasting techniques in construction industry: earned value indicators and performance models.** Machine Learning Regression Techniques (MLRT) as a shrewd method can be utilized in this study being exceptionally fruitful in demonstrating non-linear and the interrelationships among them in problems of construction projects such as the earned value indexes for tall buildings projects in Republic of Iraq. Three forecasting models were developed to fore-

### Authors' address:

Faiq M. S. Al-Zwainy  
 (<https://orcid.org/0000-0002-9948-6594>)  
 Al-Nahrain University  
 College of Engineering  
 Department of Civil Engineering  
 10072 Al-Jadryia, Baghdad, Iraq  
 e-mail: [faiq.al-zwainy@eng.nahrainuniv.edu.iq](mailto:faiq.al-zwainy@eng.nahrainuniv.edu.iq)