

## Probabilistic Time Forecasting Using Beta Distribution

Abdel Azeem S. Abdel Azeem<sup>1</sup>, Ahmed H. Ibrahim<sup>2</sup>, Hossam E. Hosny<sup>3</sup>

<sup>1</sup> Ph.D. Student, Dept. of Construction Engineering, Faculty of Engineering, Zagazig University, Egypt.

<sup>2</sup> Assistant Professor, Dept. of Construction Engineering, Faculty of Engineering, Zagazig University, Egypt.

<sup>3</sup> Professor, Dept. of Construction Engineering, Faculty of Engineering, Zagazig University, Egypt.

---

**Abstract:** The main objective of this study is to help the project managers for making better informed decisions in the form of corrective and/or preventive actions through developing a probabilistic time forecasting model, which is generated on the basis of the beta distribution as a curve fitting technique, and to provide a better basis for the schedule performance control and for the risk management of on-going projects. Four projects has been generated and used to validate the time prediction generated from the all models through the different periods of actual completion. The beta forecasting model (BFM) has been programmed in a graphical user interface (GUI) for Matlab (R2009a) and it can be implemented on all types of projects. A comparative study reveals that the BFM provides much more accurate forecasts than those are generated from the conventional methods, forecasts the completion dates on the basis of analyzing the summary of project-level data, and has accurate forecasts as the critical path method (CPM) does.

**Keywords:** Beta; Earned Value; Forecasting; Normal; Probabilistic

---

### I. Introduction

The project manager is responsible for managing his/her project and is asked to answer the following questions: (1) What are the expected completion dates and their confidence bounds? (2) What's the probability of meeting the baseline project duration? The project manager's ability, to provide a better time forecasting for his/her project in terms of the expected completion date and the probability of success, should be enhanced by a probabilistic model that provides better informed decisions in the form of corrective and/or preventive actions. The conventional deterministic methods are still deterministic and are no longer valid to provide adequate informed decisions for performance forecasting of the project. Therefore, three fundamental limitations are found in the conventional deterministic methods for schedule forecasting. First, formulas of the EVM for schedule forecasting are deterministic and provides no information about the range of possible outcomes and the probability of meeting the project objectives [1],[2]. Secondly, the schedule variance calculated from the EVM does not measure time but is expressed in a monetary unit [3]. Finally, the earned value method generates typically unreliable forecasts early in a project [4],[5].

Over the past years, different alternatives of ways have been studied to improve the schedule performance forecasting of earned value method. Vandevoorde and Vanhoucke (2006) compared three different EV-based approaches for schedule forecasting and demonstrated that the earned schedule management is the only method, among those tested methods, that provides reliable forecasting results. Kim and Reinschmidt (2010) conducted another study based on the Kalman filter algorithm and earned schedule method to forecast the estimated duration at completion (EDAC) and the probability of success. The time prediction, which is generated from both the Kalman filter forecasting method, was compared to that generated from the ES forecasting model. Such a comparison showed that the ES model generated more erratic responses to reported performance, especially in the early stages of the project, than those generated from the Kalman filter method.

Using the S-curve, as a tool to improve the schedule performance forecasting of the earned value method, has been become a major challenge for the researchers over the past years. Barraza et al. (2000; 2004) tried to use a set of S-curves generated from a network-based simulation as a project control tool and extended the concept to a probabilistic forecasting method by adjusting the parameters of probability distributions of future activities with performance indices of finished activities [6], [7]. Kim and Reinschmidt (2009) developed a probabilistic model on the basis of a beta function and Bayesian inference to forecast the estimated duration at completion. In their Bayesian approach, three parameters of the betaS-curve were derived (shape, location of mode, and project duration parameters). Useful they do, computing the posterior distributions of these parameters required a multifold integration over them, which this makes the prediction process is more sophisticated. Therefore, the BFM proposed in this study does not require a Bayesian inference and a multifold integration over the parameters as the Bayesian beta model (BBM) does. Hence, the BFM is developed only depending on a curve fitting technique such as beta function that can simulate the progress S-curve of a project efficiently.

The most common approaches for measuring the work performed are the fixed formula, weighted milestone, and percent complete techniques. Selecting the best approach during the project planning stage is based on the task duration and the number of measurement periods through its duration [8]. The fixed formula technique, where a fixed percentage of work performance is approved at the start of the work and the remaining percentage is approved at the completion of the work, is most effectively used on shorter duration tasks (shorter than two periods). While, the weighted milestone and percent complete techniques are more suitable for longer duration tasks (greater than two periods). The percent complete technique is the simplest and easiest one for the project managers to measure of the percentage of the work performed  $w_p$  for each activity, which can be calculated through dividing the quantity of work performed by the total quantity of the work. In this study, the percent complete technique is a more useful technique to estimate the earned value for each activity and provides a fair basis for evaluating the accuracy of the forecasting models through the different periods of work performed.

The objective of this paper is to introduce the BFM, which is developed on the basis of a curve fitting technique such as the cumulative beta function, and is afterward used to forecast both the expected completion dates and the probability of meeting the project duration. Then, the expected completion dates produced from the BFM and conventional deterministic forecasting methods (CDFMs) will be compared through the different periods of actual completion against those generated from the CPM, which is considered the most reliable forecasting method because it makes the time forecast at the activity-level through updating the baseline project plan periodically according to the actual performance data. Subsequently, the best time forecasting method will be selected on the basis of the lowest mean absolute percentage of error.

It should be noted that the BFM makes the time forecast at the summary of the project-level and is developed depending on the time variation in the time unit, rather than monetary one, between the PV and EV metrics and provides confidence bounds on the time predictions. Then, the confidence bounds on the time forecast can be used as an effective tool to forecast the expected completion dates and the probability of meeting the baseline project duration. This paper is organized as follows. The reviews of the conventional deterministic and the beta forecasting methods will be later discussed. In the next section, four numerical examples will be provided. Finally, the life-cycle forecasting accuracy of the all models is compared and evaluated against the time prediction generated from the CPM.

## II. Conventional Deterministic Forecasting Methods

Earned value method (EVM) was originally developed for cost management and it has not widely been used for forecasting project duration [3]. In EVM, the schedule performance index ( $SPI_{EV}$ ) can be obtained as  $EV_{AD}/PV_{AD}$  and the schedule variance ( $SV_{AD}$ ) can be obtained in monetary units as  $EV_{AD}-PV_{AD}$ . Where  $PV_{AD}$  is the planned value cost,  $EV_{AD}$  is the earned value cost, and AD is the actual data date.

At the end of a project, the  $EV_{AD} = PV_{AD} = BAC$  (budget at completion), and hence, the  $SV_{AD}$  and  $SPI_{EV}$  always equals 0 and 1 respectively. If  $SV_{AD} = 0$  and  $SPI_{EV} = 1$ , the earned value is exactly as planned, regardless of the real project status (behind, on schedule or ahead) [9]. It means that  $SV_{AD}$  and  $SPI_{EV}$  cannot give appropriate information at the late stages of the project.

Over the past years, different alternatives of ways have been studied to improve the schedule performance forecasting of the earned value method [10]; [11]; [12]; [13]. The earned schedule method (ESM) was introduced by Lipke 2003, as an extended EV metric to overcome the deficiencies of the EVM schedule indicator SPI. The planned value method (PVM) was introduced by Anbari 2003 and assumes that the time variation can be translated into time units through dividing the schedule variance by the planned value rate (PVRate), i.e.  $PVRate = BAC/PD$ . The earned duration method (EDM) was described by Jacob 2003 and extended by Jacob and Kane 2004. The earned duration ED is the product of the actual duration AD and the schedule performance index SPI, i.e.  $ED = AD * SPI$ .

The estimated duration at completion generated from these conventional deterministic methods can be calculated, on the basis of the two indicators:  $PV_{AD}$  and  $EV_{AD}$ , as defined next

$$EDAC_{EV} = PD / SPI_{EV} \dots\dots\dots (1)$$

$$EDAC_{ES} = PD / SPI_{ES} \dots\dots\dots (2)$$

$$EDAC_{PV} = PD / BAC ( BAC - SV_{AD} ) \dots\dots\dots (3)$$

$$EDAC_{ED} = PD + AD (1 - SPI_{EV}) \dots\dots\dots (4)$$

Where PD is the original estimate of the project duration, BAC is the budget cost at completion, and AD is the actual data date.

In ESM, the schedule performance index ( $SPI_{ES}$ ) can be obtained as  $ES_{AD} / AD$  and the time variation ( $TV_{AD}$ ) can be obtained in time units as  $ES_{AD} - AD$ . Then, the earned schedule ( $ES_{AD}$ ) is determined by projecting an EV onto the PV curve, as shown in Fig. 1, and then the linear interpolation between two successive planned values is generated. The term  $ES_{AD}$  is defined next:

$$ES_{AD} = k + \frac{EV_{AD} - PV_k}{PV_{k+1} - PV_k} \dots (5)$$

Where k is the longest time interval in which  $PV_k$  is less than  $EV_{AD}$ ,  $PV_k$  is the planned value at time k, and  $PV_{k+1}$  is the planned value at the next time interval, i.e. time k+1 [3]. The general framework of the earned value and the earned schedule methods is shown in Fig. 1.

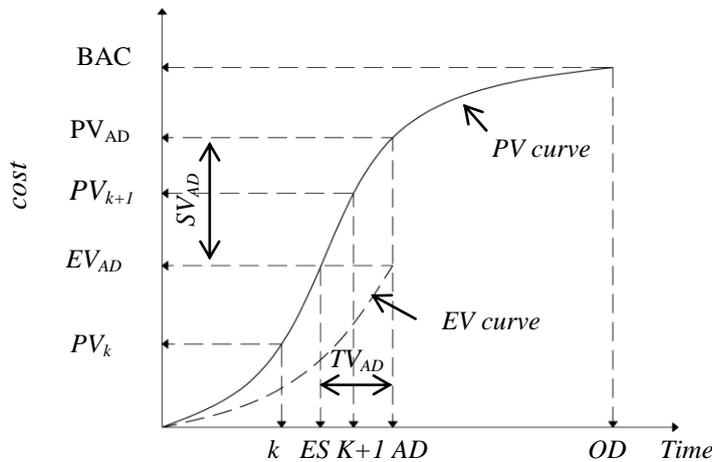


Fig. 1 : the PV and EV metrics

III. Beta Forecasting Model

The beta S-curve function is a curve fitting technique and is derived on the basis of the beta distribution, which is a continuous probability distribution defined over a finite interval 0 to 1 and is specified by the two shape parameters  $\alpha$  and  $\beta$ . The probability density function (PDF) of the beta distribution can be represented over the range of the discrete random variables %T as [14]:

$$Beta\ PDF = \frac{(\alpha - 1) * (\beta - 1) * (%T)^{\alpha - 1} * (1 - %T)^{\beta - 1}}{B(\alpha, \beta)} \dots (6)$$

Where  $B(\alpha, \beta)$  is the incomplete beta function and is defined next:

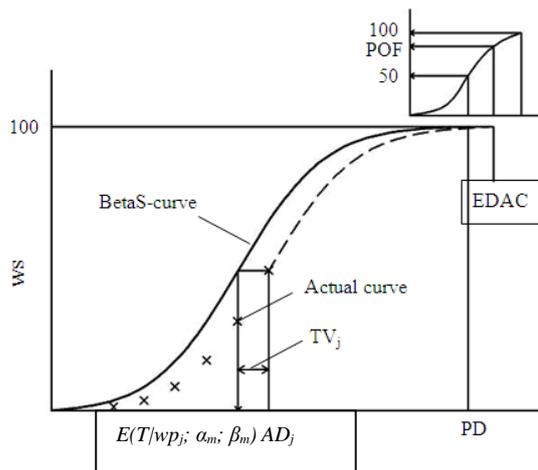
$$B(\alpha, \beta) = \int_0^1 (%T)^{\alpha - 1} * (1 - %T)^{\beta - 1} = \frac{\Gamma(\alpha) * \Gamma(\beta)}{\Gamma(\alpha + \beta)} \dots (7)$$

Where  $\alpha, \beta$  are the shape parameters;  $\alpha > 0, \beta > 0$ ; and  $\Gamma(\alpha)$  represents the gamma function. It should be noted that the minimum values of shape parameters are set to be one to confine the uni-modal shape of the beta distribution. The random variable  $%T_i$  is calculated through dividing the planned time at each measurement point by the whole project duration PD. Then, the value of each random variable  $%T_i$  can be represented over a finite interval from 0 to 1.

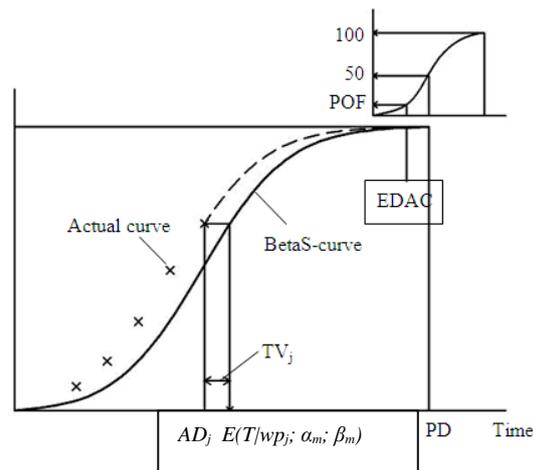
**IV. The General Framework of the Beta Forecasting Model**

The beta forecasting model is an approach that fits an S-curve function to the cumulative progress S-curve of a project acceptably. When the progress S-curve is generated from the detailed plan of the project, it is fitted by a certain model and then be used to forecast future performance of the project (Kim and Reinschmidt 2009). The general framework of the beta forecasting method is shown in Figs 2a; 2b. The horizontal axis represents the intermediate planned times and the vertical axis represents the percentage of work schedule (ws). A number of betaS-curves are generated deepening on a set of random variables for the shape parameters ( $\alpha_i; \beta_i$ ) and then evaluated in a systematic way using the least square method to select the best betaS-curve that can be therefore used to forecast the future schedule performance of a project. Then, the best betaS-curves is used to estimate the planned time for achieve each work performed  $E(T|wp_j; \alpha_m; \beta_m)$ , which can be obtained by projecting each percentage of work performed ( $wp_j$ ) on the best betaS-curve, in the light of new actual performance data. Hence, the time variation ( $TV_j$ ) can be measured, in time unit rather than monetary one, as the difference between the planned times  $E(T|wp_j; \alpha_m; \beta_m)$  and the actual data dates AD. Consequently, the time prediction generated from the BFM can be obtained by adding the  $TV_j$  to the whole project duration (PD). The dash line, as shown in Figs 2a;2b, present the extension of the actual progress curve of on-going project after the actual data date. This extension will follow the same betaS-curve to ensure that the remaining works will be done according to the detailed plans of the project.

The completion dates and their confidence bounds, which are generated from the BFM, are therefore used to estimate the probability of meeting the project duration. It should be noted that the BFM makes the time forecasts depending on the actual data, the project plan, subjective judgments, and historical data. The beta forecasting model consists of four steps: (1) Creating a library of betaS- curves, (2) Estimating the root mean square error of each betas-cure, (3) Estimating the planned time for achieving a certain work performed, (4) Forecasting the project duration, and (5) Measuring the probability of success.



**Fig. 2a:** case of behind schedule



**Fig. 2b:** case of ahead schedule

**4.1 Creating a Library of BetaS-curves:**

The planned time and planned value metrics of the project can be developed from the commercial software, such as Primavera Project Planner and Microsoft Project software, and can be used to generate the progress S-curve of the project. Then, a library of betaS-curves can be generated and then evaluated in a systematic way to find the best model that fits the progress S-curve of the project acceptably. It should be noted that generating the betaS-curves of the project using the cumulative beta function is mainly based on a set of random values for the shape parameters ( $\alpha_i; \beta_i$ ). In this study, 20,000 random variables for the shape parameters ( $\alpha_i; \beta_i$ ) are used to generate the library of betaS-curves. Note that the values of the shape parameters ( $\alpha_i; \beta_i$ ) should be positive and the values of the random variable %  $T_i$  should locate in the interval [0, 1]. The random values for the shape parameters ( $\alpha_i; \beta_i$ ) can be generated as defined next:

$$\alpha_i = \alpha_{min} + rand * (\alpha_{max} - \alpha_{min}) \dots \dots \dots (8)$$

$$\beta_i = \beta_{min} + rand * (\beta_{max} - \beta_{min}) \dots \dots \dots (9)$$

$i=1,2,3 \dots \dots \dots 20,000$

Where rand is a random number,  $(\alpha_{\min}; \beta_{\min})$  are the minimum values of shape parameters and should be one to satisfy the uni-modal condition which resembles the planned progress S-curve of the project, and  $(\alpha_{\max}; \beta_{\max})$  are the maximum values of shape parameters and are estimated to be 10. Then, the cumulative beta function, based on the random variable %T metrics and the values of the shape parameter  $(\alpha_i; \beta_i)$ , can extract the distribution of the expected values  $E(ws_i|t_i; \alpha_i; \beta_i)$  for the betaS-curves.

**4.2 Estimating the Root Mean Square Error**

20,000 models are generated using the cumulative beta function and are evaluated using the least square method that estimates the deviations, which can be measured as the differences between the expected values of a specified betaS-curve  $E(ws_i|t_i; \alpha_i; \beta_i)$  and the values of the progress S-curve ( $ws_i$ ) of the project, and minimizes the sum of the squares of them as defined next:

$$\sum_{i=1}^N E\{ [ws_i - E(ws_i|t_i; \alpha_i; \beta_i)]^2 \} = \min \dots \dots \dots (10)$$

$i=1,2,3 \dots \dots \dots N$

Where the term  $[ws_i - E(ws_i|t_i; \alpha_i; \beta_i)]$  is the deviation of a specified betaS-curve at measurement point  $i$ . The root-mean-square error (RMSE) is an indicator to measure the accuracy or strength of the generated betaS-curves in terms of closeness of fit and to provide a fair basis for the model performance evaluation later, as defined next:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N E\{ [ws_i - E(ws_i|t_i; \alpha_i; \beta_i)]^2 \}}{N}} \dots \dots (11)$$

$i=1,2,3 \dots \dots \dots N$

Where N is the number of measurement points for the progress S-curve. Then, the value of the  $RMSE_m$  and its corresponding shape parameters  $(\alpha_i; \beta_i)$  are combined in one matrix  $[RMSE_m; \alpha_i; \beta_i]$  with size  $(20,000*3)$ . Then, this matrix takes an ascending order according to the lowest  $RMSE_m$ . Hence, the values of the best-fit parameters  $(\alpha_m; \beta_m)$  can be selected as elements of the constructed matrix  $[(1, 2); (1, 3)]$ . It should be noted that the shape parameters  $(\alpha_i; \beta_i)$  refers to those the of random betaS-curve model, while the shape parameters  $(\alpha_m; \beta_m)$  refers to those of the best betaS-curve model. Table 1 indicates the best-fit shape parameters  $(\alpha_m; \beta_m)$  and their corresponding  $RMSE_m$  for the four projects. Based on such results, the values of the best-fit shape parameters  $(\alpha_m; \beta_m)$  changed between the four projects to simulate the irregular nature inherited in their progress S-curves.

**Table1:the best-fit shape parameters for the four projects**

Project	$RMSE_m$	$\alpha_m$	$\beta_m$
A	0.0220	1.186	1.402
B	0.0132	1.509	1.889
C	0.0291	1.927	2.995
D	0.0194	2.286	2.249

**4.3 Estimating the Planned Time to Achieve a Certain Work Performed**

The ultimate goal is to solicit the required planned information from the betaS-curve for the actual data. This can be acquired through obtaining the planned time required to achieve each work performed  $E(T|wp_j; \alpha_m; \beta_m)$ , which can be obtained by projecting each percentage of work performed ( $wp_j$ ) on the best betaS-curve. Therefore, the best-fit shape parameters can be efficiently used in measuring the planned time for each work performed  $E(T|wp_j; \alpha_m; \beta_m)$  at each data date. Once a project gets started, the actual progress is reported periodically and the actual data can be represented as a series of discrete values:

$$W: (wp_j, AD_j) \dots \dots \dots (12)$$

$j=1,2,3 \dots \dots \dots M$

Where  $wp_j$  is the percentage of work performed at each data date and M is the number of actual data dates points. The planned time for each work performed  $E(T|wp_j; \alpha_m; \beta_m)$  at each data date can be measured as shown in Fig. 2 and defined next:

$$E(T|wp_j; \alpha_m; \beta_m) = beta^{-1}(wp_j/100; \alpha_m; \beta_m) * PD \dots \dots \dots (13)$$

$j=1,2,3 \dots \dots \dots M$

Where  $\beta^{-1}$  represents the inverse of the cumulative beta function and computes the values  $E(T|wp_j; \alpha_m; \beta_m)$  in the light of new actual performance data  $wp_j$ . Note that the term  $\beta^{-1}(wp_j/100, \alpha_m, \beta_m)$  is equivalent to the random variable %T and is obtained as a percentage from the planned project duration PD.

**4.4 Forecasting the Project Duration**

The beta forecasting method is developed mainly on the basis of time variation in the time dimension between the PV and EV curves. The time variation  $TV_j$  can be measured, in time unit rather in monetary one, as the differences between the planned times  $E(T|wp_j; \alpha_m; \beta_m)$  and actual data dates  $AD_j$ , as defined next

$$TV_j = AD_j - E(T|wp_j; \alpha_m; \beta_m) \dots \dots \dots (14)$$

Consequently, the time prediction generated from the BFM can be calculated basically from adding the time variation  $TV_j$  to the planned project duration PD. The  $EDAC_j$  observed from the BFM can be computed through the different points of actual completion, as defined next

$$EDAC_j = PD + TV_j \dots \dots \dots (15)$$

**4.5 Measuring the Probability of Meeting the Project Duration:**

The probabilistic model BFM, in distinction to the conventional deterministic forecasting methods, provides prediction bounds of project duration and probability of its success profiles, which collectively or separately provide an early warning point determined by the project management according to a predetermined acceptable risk level. Statistical probability function such as cumulative normal distribution can be used to estimate the probability of success. The standard deviation of the project duration ( $\sigma_T$ ), which is an important parameter for the cumulative normal distribution, can be estimated on the basis of historical data and subjective judgments. In this study, three point estimate (PERT estimate) was used to evaluate the standard deviation of the project duration  $\sigma_T$ . For example, project A, the value of the standard deviation of the project duration  $\sigma_T$  can be directly calculated on the basis of the pessimistic ( $P=1.40 * PD$ ) and optimistic ( $O=0.80 * PD$ ) durations.

$$\sigma_T = \frac{(P-O)}{6} = \frac{1.40*70-0.80*70}{6} = 7 \text{ weeks} \dots \dots \dots (16)$$

Then, the  $POF_j$  of the project can be directly measured in terms the two parameters (PD;  $\sigma_T$ ) of the cumulative normal distribution and the  $EDAC_j$ . For example, for the project A, the  $POF_j$  can be directly as  $\text{norm}^{-1}(83.68 \text{ weeks}, 70 \text{ weeks}, 7 \text{ weeks})$ , where  $\text{norm}^{-1}$  is the inverse of the cumulative normal function, 83.68 weeks is the EDAC at the end of period no. 17<sup>th</sup>, and 7 weeks is the standard deviation of the project duration  $\sigma_T$ . In this case, the project A is behind schedule because the  $EDAC_j = 83.68 \text{ weeks}$  is greater than  $PD = 70 \text{ weeks}$ . In case of behind schedule, the elements contributed in Fig. 2a will be:  $E(T|wp_j; \alpha_m; \beta_m) \leq AD_j$ ,  $TV_j \geq 0$ ,  $EDAC_j \geq PD$ , and  $POF_j \geq 0.50$ . In case of ahead schedule, on the other hand, the elements contributed in Fig. 2b will be:  $E(T|wp_j; \alpha_m; \beta_m) \geq AD_j$ ,  $TV_j \leq 0$ ,  $EDAC_j \leq PD$ , and  $POF_j \leq 0.50$ . Once, the probability of failure  $POF_j$  is measured, and then the probability of success  $POS_j$  will be  $1-POF_j$ .

**V. Numerical Examples**

The BFM model formulated in the previous sections has been programmed in a graphical user interface (GUI) of Matlab (R2009a). The GUI of Matlab is to make the programming of PFM easier for the interested users and project managers. The required inputs to run the BFM are the  $PV_{AD}$ ,  $EV_{AD}$ , and  $AD$ . The possible outputs of the probabilistic BFM are the  $EDAC_j$  and the probability of success  $POS$ , i.e. the probability of meeting the project duration. On the other hand, the possible outputs of the deterministic methods CDFMs are the expected project duration and estimated cost at completion. Moreover, the PFM provides prediction bounds for EDAC, three point estimate, lower bound (LB), mean (M), upper bound (UP). The CDFMs, on the other hand, don't provide prediction bounds of EDAC forecasting. four artificial examples (A, B, C, D, and E) are presented in Table 2 to overcome limited and often incomplete real project data and to validate the results observed from all the models.

**Table 2: the data of the four artificial projects**

Project	BAC (EGP.)	PD (weeks)	Number of activities	Number of actual observations
A	2,290,000	70	18	20
B	1,850,000	76	22	18
C	2,040,000	48	19	14
D	2,750,000	100	13	20

The planned value cost  $PV_{AD}$  at each data date can be directly computed from the baseline Gantt chart. Hence, the percentage of schedule complete (ws) at each data date can be determined through dividing the planned value cost  $PV_{AD}$  by the budget at completion cost (BAC), i.e.  $ws = PV_{AD} / BAC$ . Once, a project gets started, and then each percentage of work performed wp for each activity can be measured from the construction site. Then, the earned value cost for each activity can be calculated on the basis of the percent complete technique, which is previously explained in the literature review, as the product of the percentage of work performed wp and the budget cost for each activity, i.e.  $EV_{AD} = wp * \text{budget cost for each activity}$ . The cumulative earned value cost  $EV_{AD}$  is therefore calculated by summation of the earned value cost for each activity. The estimated duration at completion of CPM can be directly calculated through updating process for the durations and dates of the remaining activities, based on the actual reporting data. However, the time prediction produced from the PFM and the other CDFMs will be assessed and compared from the start to the end of the project to that produced from the CPM.

The status of project A is ahead schedule at the early stages and behind schedule in the late stages. On the other hand, the status of project B contradicts the status of Project A. On the other hand, the statuses of the two Projects C and D during the whole duration are behind and ahead schedule respectively. These changes in the projects statuses during the different projects periods of actual completion are intended to assess the response of all the models to the periodic changes in the actual performance data.

### VI. Comparative Study

The time prediction generated from the all models are compared and assessed against that generated from the CPM, which is the most reliable method because it makes the time forecast at the activity-level through the updating process. The absolute percentage of error ( $APE_j$ ) and mean absolute percentage of error (MAPE) are then measured as:

$$APE_j = \frac{100}{EDAC_{cpm}} * \left| EDAC_j - EDAC_{cpm} \right| \dots\dots(17)$$

$$MAPE = \frac{100}{M} * \sum_{j=1}^M APE_j \dots\dots\dots(18)$$

$j=1,2,3,\dots\dots\dots M$

Where  $EDAC_j$  is the estimated duration at completion that is generated from the all models, M is the number of the actual data dates, and  $EDAC_{cpm}$  is the estimated duration at completion that is generated from the CPM. The accuracies of all models were assessed and compared through the different project periods of work performed wp. Table 3 represents a comparative study between the all models, for the four projects along their construction lifecycles, and provides the values of the APE during the different periods of completion. The columns ( $APE_{EV}$ ;  $APE_{ES}$ ;  $APE_{PV}$ ;  $APE_{ED}$ ) indicated the APE results generated from CDFMs for the four projects, while column ( $APE_{BF}$ ) indicated the APE profile produced from the BFM. The absolute percentage of error  $APE_j$  of all the models can be measured at the different evaluation periods of the performed work percentage wp. These evaluation periods are estimated to 0-10% BAC, 10-20% BAC, 20-30% BAC, 30-40% BAC, 40-50% BAC, 50-60% BAC, 60-70% BAC, 70-80% BAC, 80-90% BAC, and 90-100% BAC. It should be noted that the allowable error for all the models is 5%. If the APE value of a model is greater than the allowable error, the cell is then filled with the yellow color to demonstrate larger errors generated from a specified model. In other words, the time prediction of a specified model can be accepted if its APE is less than 5%.

Based on such comparative study, the results of APE reveal that the BFM is the only method that shows accurate forecasts during the different periods of actual completion. However, the earned value and the earned schedule methods generate more erratic responses of time forecasting at the early stages of all projects. As more actual data reported, they generate acceptable responses of time prediction to that is generated from the CPM. On the other hand, the planned value and the earned duration methods generates unreliable responses of time forecasting at the end of the projects A, C, and D. All CDFMS, however, provides no information about the range of possible outcomes and the probability of meeting the project objectives. The major merit of the probabilistic model BFM, in distinction to conventional deterministic forecasting methods, is that it generates more acceptable responses for the time prediction and has accurate forecasts during all different periods of the actual completion for the four projects.

### VII. Comparative Study Summary

The mean absolute percentage of error MAPE generated from the BFM and CDFMs were calculated and presented in Table 4. Based on such MAPE results, the beta forecasting approach is the only method which shows reliable time forecasting during the whole project duration and can predict the project duration with

smaller errors (MAPE is less than 5%). The BFM can forecast the project duration, with MAPE less than 5%, indicating that a higher accuracy of time forecast can be obtained from the BFM at the summary of project-level without analyzing the activity-level technical data as the CPM does. Regarding to the CDFMs, the earned value and earned schedule showed more erratic responses for the time forecasting, especially at the early stages of all projects. The planned value and the earned duration methods generated unreliable responses of time forecasting at the end of the projects. Therefore, the BFM provides a reliable estimating for the progress S-curve, which in turn can help the project managers for making informed decisions regarding to the schedule performance of the project. Moreover, the BFM provides confidence bounds on predictions, which can be used to determine the range of probable project duration at completion and the probability of its success.

**Table 3: APE of the all models for all projects**

Project	Error	0-10%	10-20%	20-30%	30-40%	40-50%	50-60%	60-70%	70-80%	80-90%	90-100%
A	APE <sub>BF</sub>	2.02%	0.63%	0.92%	0.90%	1.16%	0.39%	1.64%	3.96%	2.23%	0.32%
	APE <sub>EV</sub>	28.18%	8.31%	8.31%	6.72%	0.69%	1.50%	2.83%	1.41%	3.73%	9.66%
	APE <sub>ES</sub>	18.02%	9.32%	6.35%	10.06%	1.26%	1.02%	0.76%	2.16%	4.57%	1.20%
	APE <sub>PV</sub>	0.60%	1.16%	0.77%	0.37%	0.81%	1.57%	1.06%	2.13%	5.28%	9.98%
	APE <sub>ED</sub>	0.39%	1.50%	0.40%	1.15%	0.93%	1.90%	1.86%	3.04%	5.58%	9.78%
B	APE <sub>BF</sub>	0.80%	1.55%	1.26%	0.19%	2.08%	1.02%	0.84%	1.02%	0.68%	0.43%
	APE <sub>EV</sub>	36.71%	26.63%	9.41%	17.50%	5.28%	4.77%	0.00%	0.00%	0.78%	2.66%
	APE <sub>ES</sub>	23.15%	15.27%	8.91%	6.29%	5.10%	5.40%	0.00%	0.00%	0.80%	0.61%
	APE <sub>PV</sub>	1.14%	0.36%	0.16%	4.81%	1.12%	1.77%	0.00%	0.00%	1.04%	2.73%
	APE <sub>ED</sub>	0.25%	0.53%	0.13%	3.56%	0.49%	1.23%	0.00%	0.00%	1.29%	2.89%
C	APE <sub>BF</sub>	2.16%	5.00%	0.34%	0.46%	-----	1.80%	1.64%	1.81%	2.03%	2.77%
	APE <sub>EV</sub>	126.5%	184.4%	80.65%	44.52%	-----	22.30%	6.22%	0.56%	3.88%	10.32%
	APE <sub>ES</sub>	67.36%	53.73%	29.63%	20.43%	-----	16.15%	9.76%	6.19%	8.04%	3.25%
	APE <sub>PV</sub>	2.25%	13.01%	11.57%	11.57%	-----	9.80%	1.39%	2.83%	5.23%	10.48%
	APE <sub>ED</sub>	3.47%	11.25%	6.45%	4.84%	-----	4.39%	1.95%	5.05%	6.69%	10.59%
D	APE <sub>BF</sub>	3.32%	1.78%	2.02%	0.46%	2.98%	2.61%	1.08%	0.95%	0.73%	3.26%
	APE <sub>EV</sub>	29.37%	27.79%	40.27%	30.29%	26.77%	27.85%	15.33%	9.86%	2.19%	9.77%
	APE <sub>ES</sub>	28.13%	24.95%	18.74%	12.81%	7.86%	7.54%	4.78%	6.98%	5.06%	1.60%
	APE <sub>PV</sub>	1.76%	3.95%	3.16%	4.48%	6.20%	10.36%	5.78%	4.44%	1.41%	10.23%
	APE <sub>ED</sub>	0.28%	1.10%	14.84%	11.62%	11.44%	15.01%	5.77%	2.60%	2.80%	11.94%

**Table 4: MAPE of all models**

Method	Projects			
	A	B	C	D
BFM	1.54%	0.91%	2.16%	2.24%
EVM	5.53%	12.07%	48.60%	21.17%
ESM	4.21%	7.51%	23.14%	12.29%
PVM	3.10%	1.49%	7.49%	5.24%
EDM	3.41%	1.26%	6.92%	7.50%

**VIII. Estimated Duration at Completion Profile**

The probabilistic EDAC profile generated from the BFM consists of three curves: M (mean of project duration), UB (upper bound), and LB (lower bound) curves. When the three curves combined together, they show the range of possible completion dates at a given confidence level at each data date [2]. The expected completion dates and their prediction bounds, which are generated from the BFM, represented the history of probabilistic predictions for the project duration from the start of a project to the point of forecasting. The confidence limits (bounds) of the EDAC can be directly generated from the BFM in terms of the standard deviation of the project duration ( $\sigma_T$ ) and a specified confidence interval selected by the user. The project manager is asked to select an appropriate confidence level on the basis of historical data and subjective judgments. The prediction bounds are determined at the confidence level selected by the user. Note that a confidence level 90% was used in this study. This means that upper and lower bounds are determined at the 5% confidence level on each side. Therefore, the confidence bounds have 90% probability of including the actual project duration.

The thick black solid line, as shown in Figs. 3a, 4a, 5a, and 6a, represents the EDAC profile generated from the BFM. Moreover the BFM provides confidence limits for the completion dates, which in turn can help the project managers to make better informed decisions in the form of corrective and/or preventive actions to maintain and/or restore the baseline schedule.

**IX. Probability of Success Profile**

The probability of success profile which is defined as the changes in the probability of meeting the project objective (planned duration). This profile also can be used to detect an early warning point at a specific risk level [2]. It should be noted that the POS profile is inversely proportional with the EDAC profile. Literally, if the EDAC profile increases, the POS profile tends to decrease. The decrease in the POS profile, for example, between the end of reporting periods no. 8<sup>th</sup> and 10<sup>th</sup> as shown Figs. 4a and 4b, is corresponding to the increase in the EDAC profile at the same periods. This is because the inverse proportional between the EDAC and POS profiles. The POS and EDAC profiles can be collectively used to judge and forecast the schedule performance through identifying the current status and time forecast for the projects at each actual data date. The next sections reviews the current status and time forecast for the four projects.

**Project A**

The status of the Project A, as shown in Fig. 3b, has been ahead schedule by the end of reporting period no. 7<sup>th</sup> and has been thereafter changed to be behind schedule. The probability of success for the Project A has improved from 50% at the beginning of the project to reach 85% at the end of reporting period no. 5<sup>th</sup>. Thereafter, the probability of success has declined gradually to reach zero at the end of reporting period no. 20<sup>th</sup> with no chance of meeting the project duration (70 weeks). Then, the Project A has finished with a  $POS_j = 0\%$  and current  $EDAC_j = 80$  weeks (delayed 10 weeks).

Based on the results above, some conclusions can be drawn about the future schedule performance of the Project A. Both the EDAC and the probability of success profiles, as shown in Figs. 3a and 3b, indicated that an early warning of a project overrun should have been issued at the reporting period no. 9<sup>th</sup> at a 5% risk level. Note that if the LB of EDAC profile has exceeded the original duration line (70 week), this means that the project cannot meet its objective and should be achieved with behind schedule status. Considering that this project has 52% complete at the reporting period no. 9<sup>th</sup>, thereafter, finishing the project on its original duration is no longer a realistic objective and the plan should be revised according to the current information available. Hence, the BFM helps the project managers to generate early warnings to take corrective action to restore the schedule of a project.

**Project B**

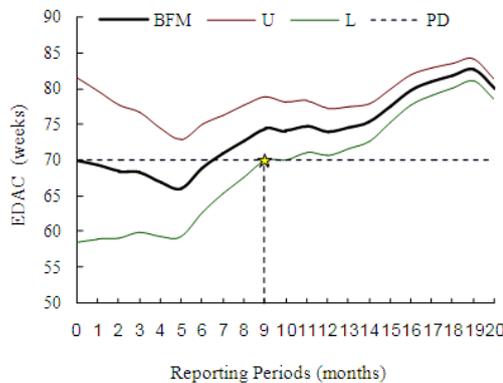
The status of the Project B, as shown Figs. 4a and 4b, has been behind schedule by the end of reporting periods no. 11<sup>th</sup> and thereafter changed to be ahead schedule. Then, the final status of project B is ahead schedule with 99% chance of finishing on its planned duration (76 weeks).

**Project C**

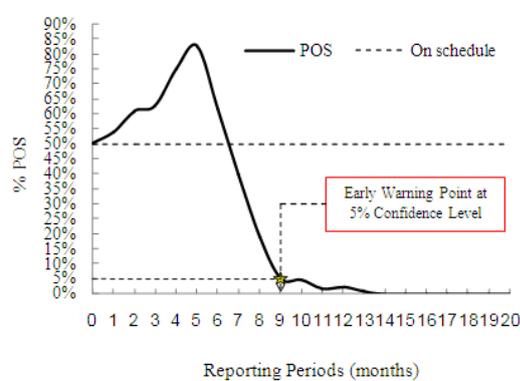
The status of the Project C, as shown in Figs. 5a and 5b, has been behind schedule along the whole duration. An early warning of behind schedule for the Project C should have been issued at the end of reporting period no. 3<sup>th</sup> at a 5% risk level. Considering that this project has 5% complete at the reporting period no. 3<sup>th</sup>, thereafter, finishing the project on its original duration is no longer a realistic objective and the plan should be revised according to the current information available. Then, the Project C has finished, has behind schedule status, and has no chance of finishing on its planned duration (48 weeks).

**Project D**

The status of the Project D, as shown in Figs. 6a and 6b, has been ahead schedule along the whole duration. Then, the final project status is ahead schedule with 99% chance of finishing on its planned duration (100 weeks). Then, the Project D has finished, has ahead schedule status, and has 99% chance of finishing on its planned duration (100 weeks).



**Fig. 3a:** EDAC Profile for Project A



**Fig. 3b:** POS Profile for Project A

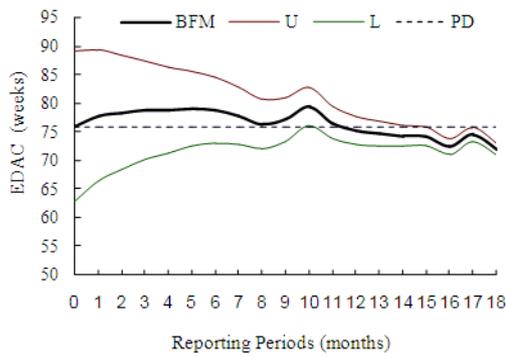


Fig. 4a: EDAC profile for Project B

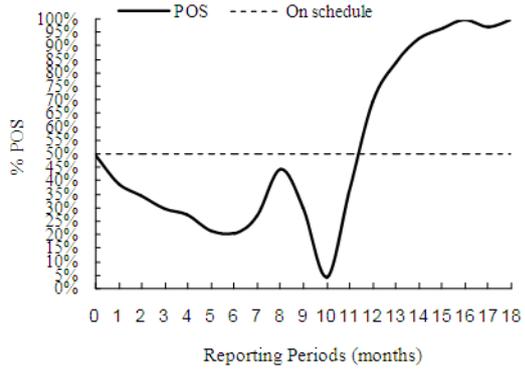


Fig. 4b: EDAC profile for Project B

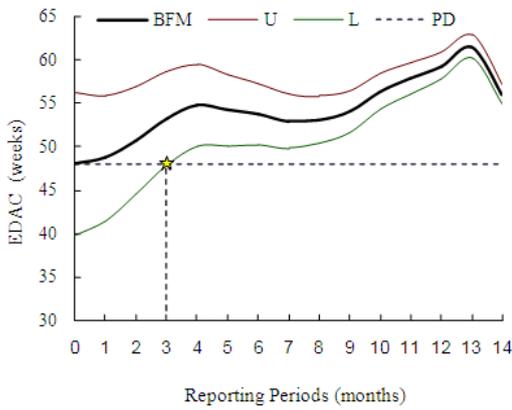


Fig. 5b: EDAC profile for Project C

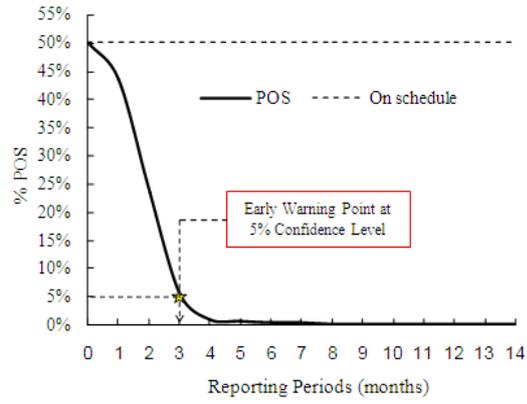


Fig. 5a: EDAC profile for Project C

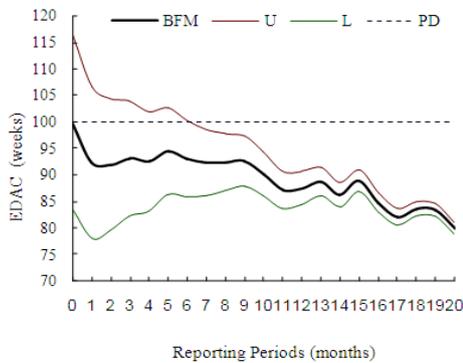


Fig. 6b: EDAC profile for Project D

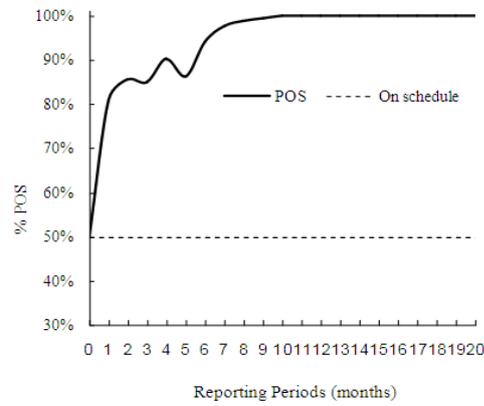


Fig. 6a: EDAC profile for Project D

### X. Comparisons to the Previous Studies

The results drawn from this study were compared to the previously found results that were developed by Vandevorde and Vanhoucke (2006) and Kim and Reinschmidt (2009). Vandevorde and Vanhoucke (2006) concluded that the ESM is the best schedule forecasting method. Their results and conclusions for the ESM were applied for the projects that have a short lifecycle, neglecting the negative effect of linear interpolation inherited in its formula (stated in Equ. 5) for schedule forecasting. Fig. 7, which is developed by Kim and Reinschmidt 2009, shows the MAPE of the BBM and ESM obtained at different evaluation points. These evaluation points were estimated to be 5%PD, 40%PD, 60%PD, 80%PD, and 90%PD. Their study demonstrated that the probabilistic BBM can be applied through the different periods of actual completion with smaller errors (MAPE is less than 5%) than those observed in ES method. Based on such the results, the earned schedule model produced more significant errors, particularly in the beginning periods of the project than the BBM.

Four projects, which their baseline durations vary from 12 to 25 months and have 72 actual observations, are generated independently and are enough to draw a statistically meaningful conclusion, and are

used to obtain the APE at different stages of the construction project life cycle. In this study, the conclusion drawn for the ESM is that it produces larger errors for time prediction in the early stages of the projects, contradicts the previously results generated from Vandevoorde and Vanhoucke (2006), and approves the previously results generated from Kim and Reinschmidt (2009). On the other hand, the results generated from the planned value and the earned duration methods confirms the previously found results of Vandevoorde and Vanhoucke (2006).

The BFM provides sufficiently accurate forecasts: MAPE less than 5% even very early in the project. However, it should be noted that the BFM does not require a Bayesian inference and multifold integration over the parameters as the Bayesian beta model (BBM) does and does not rely on activity level performance data and analysis as the CPM does.

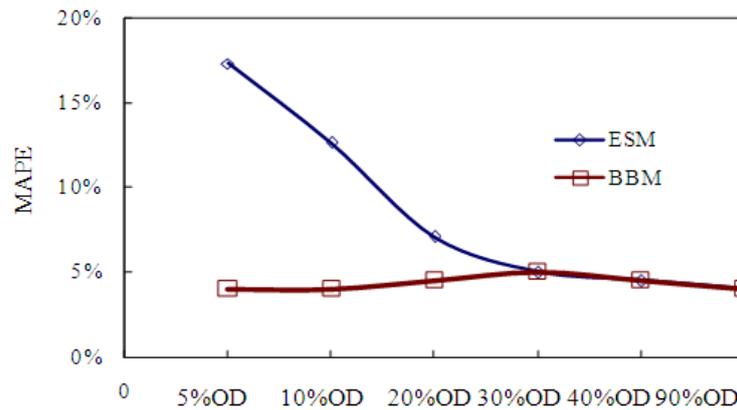


Fig. 7: MAPE at different stages of completion [1]

### XI. Conclusions and Recommendations

The main objective of this paper is to develop a probabilistic method for the time forecasting, on the basis of beta function as a curve fitting technique. This probabilistic method is used to answer the following questions: (1) What are the expected completion dates and their confidence bounds? (2) What’s the probability of meeting the baseline project duration? (3) Is the overall schedule performance improving or deteriorating? Another objective is to compare and assess the time forecasting, generated from both the BFM and the conventional deterministic forecasting methods, against that generated from the critical path method. It should be noted that the proposed method BFM does not require a Bayesian inference and multifold integration over the parameters as the method developed by Kim and Reinschmidt 2009.

A comparative study reveals that the BFM provides much more accurate forecasts than the conventional deterministic forecasting methods (CDFMs) and as accurate forecasts as the critical path method without analyzing activity-level technical data. The major merit of the probabilistic model BFM, in distinction to conventional deterministic forecasting methods, is that it provides prediction bounds of project duration at completion and probability of its success profiles, which in turn can help project managers to make better informed decisions in the form of corrective and/or preventive actions, and can be applied at the different periods of actual completion of a project with smaller errors than those observed in the deterministic forecasting methods.

The BFM is applied to four projects which their baseline durations vary from 12 to 25 months and 72 actual observations. The aim of these projects is to demonstrate how effectively the BFM can be used to improve the accuracy of project forecasts, improve upon deterministic forecasting methods by adaptively developing the forecasts of project duration, by developing quantitative confidence intervals around these forecasts, and by providing project managers with early warnings of possible project overruns. That is, probabilistic outputs from the BFM are visualized with two schedule risk charts (Figs. 3 to 6): the probabilistic EDAC and the probability of success profiles. Used separately or in combination, these charts can serve as a time-saving project management dashboard, which is ideal for real-time monitoring of the overall schedule risk from various perspectives, for example: (1) What is the probability of finishing the project on time? (2) Is the overall schedule performance improving or deteriorating? Most of all, project managers using the PFM will be able to identify when projects are in control and when attention is needed long before the project has deteriorated in a quantitative way. The confidence level for detecting such an early warning point can be determined by the project management according to a predetermined acceptable risk level for the project.

The conclusions drawn from this study for the deterministic methods were compared to the previously found conclusions of Vandevoorde and Vanhoucke (2006) and Kim and Reinschmidt (2009). Our results generated from the earned schedule method confirms the previously found results of Kim and Reinschmidt (2009) and contradicts those of Vandevoorde and Vanhoucke (2006). On the other hand, our results generated from the planned value and the earned duration methods confirms the previously found results of Vandevoorde and Vanhoucke (2006).

The BFM has been programmed in a graphical user interface (GUI) for Matlab (R2009a) and it can be implemented on all kinds of projects. The software of the BFM program is available on request by the authors. A limitation of the BFM is that it is applicable only to the prediction of project duration at completion and probability of success of meeting the project duration, not to the prediction of project cost at completion (EAC). Due to space limitations, this research cannot be reported here but will be covered in a subsequent publication. However, the beta functions can be extended to forecast of the cost at completion so that schedule and cost forecasting can be integrated within a consistent methodology. Finally, it has to be noted that the work described in this paper is a part of a Ph.D. dissertation (incomplete thesis) that has been currently conducted [15].

### References

- [1]. Kim, B. C.; and Reinschmidt, K. F. "Probabilistic Forecasting of Project Duration Using Bayesian Inference and the Beta Distribution." *J. Constr. Eng. Manage.*, 135(3), 2009, 178–186.
- [2]. Kim, B. C.; and Reinschmidt, K. F. "Probabilistic Forecasting of Project Duration Using Kalman Filter and the Earned Value Method." *J. Constr. Eng. Manage.*, 130(1), 2010, 25–32.
- [3]. Vandevoorde, S.; and Vanhoucke, M. "A Comparison of Different Project Duration Forecasting Methods Using Earned Value Metrics." *Int. J. Proj. Manage.*, 24(4), 2006, 289–302.
- [4]. Christensen, D. S.; and Heise, S. R. "Cost Performance Index Stability." *National Contract Management Journal*, 25 (1), 1993, 7–15.
- [5]. Zwikael, O.; Globerson, S.; and Raz, T. "Evaluation of Models for Forecasting the Final Cost of a Project." *Proj. Manage. J.*, 31 (1), 2000, 53–57.
- [6]. Barraza, G. A.; Back, W. E.; and Mata, F. "Probabilistic Monitoring of Project Performance Using SS-curves." *J. Constr. Eng. Manage.*, 126(2), 2000, 142–148.
- [7]. Barraza, G. A.; Back, W. E.; and Mata, F. "Probabilistic Forecasting of Project Performance Using Stochastic S Curves." *J. Constr. Eng. Manage.*, 130(1), 2004, 25–32.
- [8]. PMI. "Practice Standard for Earned Value Management." *Proj. Manage. Institute, Inc.*, Newtown Square, Pa, 2005.
- [9]. Vanhoucke, M. "Earned Value Management." *Proj. Manage. with Dynamic Scheduling*, Part 3, 2012, 215-238.
- [10]. Lipke, W. H. "Schedule is Different." *The Measurable News*, Summer, 2003, 31–34.
- [11]. Anbari, F. "Earned Value Method and Extensions." *Proj. Manage. J.*, 34(4), 2003, 12–23.
- [12]. Jacob, D. "Forecasting Project Schedule Completion with Earned Value Metrics." *The Measurable News*, Summer, 2003, 7–9.
- [13]. Jacob D. S. and Kane, M. "Forecasting Schedule Completion Using Earned Value Metrics Revisited." *The Measurable News*, Summer, 2004, 11–7.
- [14]. Matlab. "User Guides of Matlab Software." *MathWorks*, Version R2009a, License Number 161051.
- [15]. Abdel Azeem S.A. "Forecasting of Project Cost and Schedule Performance Using Probabilistic and Deterministic Models." Ph.D. Dissertation ( Incomplete Thesis ), Faculty of Engineering, Zagazig University, Egypt.