POLITECNICO DI TORINO
Repository ISTITUZIONALE

Combination of Growth Model and Earned Schedule to Forecast Project Cost at Completion

Original
Combination of Growth Model and Earned Schedule to Forecast Project Cost at Completion / Narbaev, T., DE MARCO, A.. - In: JOURNAL OF CONSTRUCTION ENGINEERING AND MANAGEMENT. - ISSN 0733-9364. - STAMPA. - 140:1(2014), p. Article number 04013038. [10.1061/(ASCE)CO.1943-7862.0000783]

Availability:
This version is available at: 11583/2517098 since:

Publisher:
ASCE - American Society of Civil Engineering

Published
DOI:10.1061/(ASCE)CO.1943-7862.0000783

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

18 November 2026

Combination of Growth Model and Earned Schedule to Forecast Project Cost at Completion

Timur Narbaev, Ph.D.[1]; and Alberto De Marco, Ph.D.[2]

Abstract: To improve the accuracy of early forecasting the final cost at completion of an ongoing construction project, a new regression-based nonlinear cost estimate at completion (CEAC) methodology is proposed that integrates a growth model with earned schedule (ES) concepts. The methodology provides CEAC computations for project early-stage and middle-stage completion. To this end, this paper establishes three primary objectives, as follows: (1) develop a new formula based on integration of the ES method and four candidate growth models (logistic, Gompertz, Bass, and Weibull), (2) validate the new methodology through its application to nine past projects, and (3) select the equation with the best-performing growth model through testing their statistical validity and comparing the accuracy of their CEAC estimates. Based on statistical validity analysis of the four growth models and comparison of CEAC errors, the CEAC formula based on the Gompertz model is better-fitting and generates more accurate final-cost estimates than those computed by using the other three models and the index-based method. The proposed methodology is a theoretical contribution towards the combination of earned-value metrics with regression-based studies. It also brings practical implications associated with usage of a viable and accurate forecasting technique that considers the schedule impact as a determinant factor of cost behavior. **DOI: 10.1061/(ASCE)CO.1943-7862.0000783.**

Author keywords: Construction management; Cost forecasting; Earned schedule; Earned value management; Growth model; Nonlinear regression; Cost and schedule.

Introduction

With the purpose of controlling cost overrun and schedule delays, earned-value management (EVM) is often used as an objective technique for supporting the tasks of monitoring, analyzing, and forecasting project cost and schedule performance. Earned-value management is a method integrating a project's cost, schedule, and scope metrics into a single measurement system (PMI 2008). It indicates that cost deviations may affect the schedule progress and vice versa, i.e., a project that is lagging behind or is ahead of schedule is more likely to experience changes in the cost plan.

In particular, EVM is a widely accepted method to compute cost estimate at completion (CEAC) of an ongoing project based on current progress and performance. Major contributions to EVM-based CEAC were established in the 1990s within U.S. defense projects. Comparatively, the construction industry experienced small applications of these findings and during the past decade little advancement in CEAC forecasting research for construction projects has been reported (De Marco and Narbaev 2013).

Although credited to be simple-to-use and widely accepted formulas, index-based methods for CEAC forecasting have three primary limitations, as follows: (1) reliance on past cost performance only, (2) unreliable forecasting in early stages of a project life, and (3) no count of forecasting statistics (Fleming and Koppelman 2006; Kim and Reinschimdt 2010; Tracy 2005; Zwikael et al. 2000). These three limitations are the primary reasons to further extend the application boundaries of index-based methods through their refinement or integration with other forecasting techniques.

The most reported alternative techniques to index-based cost forecasting formulas are those that use linear or nonlinear regression analysis to develop regression-based models (Christensen et al. 1995). These methods are regarded as more sophisticated in calculating a project's CEAC than index-based methodologies but are able to generate better estimates early in a project life (Tracy 2005). Nonlinear formulas better describe nonlinear relationships between input and output variables and are frequently used to build the nonlinear cost growth pattern. Furthermore, the s-shaped curve of this cumulative cost is produced by the sigmoid models, also termed growth models. These methodologies extend the application boundaries of the traditional index-based CEAC methods and overcome the three previously noted limitations. In the literature, growth models with nonlinear regression have been widely applied to study cumulative cost growth (Christensen et al. 1995).

Although some literature is available in the respective areas of application of either index-based or regression-based models, very little research has been carried out to combine these two techniques for calculating the CEAC of ongoing construction projects. In an attempt to refine CEAC methodologies and improve their accuracy, this paper contributes to the extension of the EVM body-of-knowledge by filling this research gap through the integration of EVM methods into growth models.

To this end, a nonlinear regression-based CEAC methodology is proposed to interpolate characteristics of growth models and

[1]Assistant Professor, International School of Economics, Kazakh-British Technical Univ., 59 Tole bi St., Almaty 050000, Kazakhstan; formerly, Research Fellow, Dept. of Management and Production Engineering, Politecnico di Torino, 10129 Torino, Italy (corresponding author). E-mail: t.narbaev@kbtu.kz

[2]Assistant Professor, Dept. of Management and Production Engineering, Politecnico di Torino, corso Duca degli Abruzzi 24, 10129 Torino, Italy. E-mail: alberto.demarco@polito.it

combine the earned schedule (ES) concept into an equation to compute a construction project's CEAC. Earned schedule is an effective EVM metric that helps analyze the schedule progress and estimate a project's completion time.

Three primary tasks were carried out in the research that is reported in this paper to achieve the previously noted purpose, as follows: (1) a new CEAC formula was developed with integrated application of four candidate growth models and ES concepts, (2) this formula was validated through testing its applicability on construction case projects, and (3) the statistically valid growth model was selected for further CEAC comparative analysis with the traditional index-based method. The prediction accuracy of the four model equations is based on a comparison of their estimate errors.

This paper is structured as described next. First, the writers briefly describe the index-based and regression-based cost-forecasting methods pertinent to the research reported in this paper, and bring the four growth models and the ES method into the context. Second, the proposed methodology is presented through a stepped procedure that formulates the growth models, develops the CEAC equations, and integrates ES. The technique is then validated on nine construction projects and the models' CEAC estimate errors are compared with results discussions. The paper then presents implications, limitations, future research directions, and conclusions.

Background Research

Earned-Value Management Cost-Forecasting with an Index-Based Method

Earned-value management is a technique of thoroughly quantifying the technical performance of an ongoing project and integrating it with cost and time. It is a powerful tool that allows objective monitoring of actual status, and comparing it with a plan, tracking deviations from the project baseline, and forecasting the final cost and time at completion based on the current project status (PMI 2008). The key parameters representing fundamentals of analysis are the planned value (PV), earned value (EV), actual cost (AC), and budget at completion (BAC). Fig. 1 is a graphical illustration of these metrics and exposes the standard condition of a construction project, i.e., over-budget and behind schedule. These four metrics together with the associated cost performance index (CPI = EV/AC, indicating how efficiently a project is using its resources) and schedule performance index (SPI = EV/PV, showing how effectively time is spent) are used to analyze a project's cost and schedule status, and provide forecasts of project cost and time at completion.

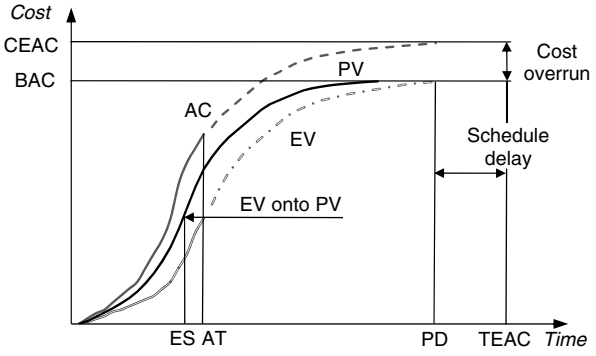


Fig. 1. Earned-value management and ES metrics

Cost estimate at completion forecasting and time estimate at completion (TEAC) forecasting are performed by extrapolating the actual project cost performance and schedule progress to the end of the project. The Project Management Institute (PMI 2011) provides two widely accepted formulas to calculate these two forecasts. For this, the original values of BAC and planned duration (PD) are corrected by the corresponding indices to take into account past performance and progress, respectively, as in Eqs. (1) and (2) (PMI 2011):

$$\text{CEAC} = \text{AC} + \frac{\text{BAC} - \text{EV}}{\text{CPI}} = \frac{\text{BAC}}{\text{CPI}} \tag{1}$$

$$\text{TEAC} = \left(\frac{\text{BAC}}{\text{SPI}}\right) \Big/ \left(\frac{\text{BAC}}{\text{PD}}\right) = \frac{\text{PD}}{\text{SPI}} \tag{2}$$

One of the major findings in cost-forecasting is that the value of the final CPI-based estimate does not vary by more than 10% from its value at 20% completion and after that point it tends to worsen (Christensen and Heise 1993). However, recent studies show these are applicable only to large and long-duration defense projects (Lipke et al. 2009).

Even though CEAC forecasting with Eq. (1) is widely used as a standard (Anbari 2003; Fleming and Koppelman 2006; PMI 2011), its limitations have been largely reviewed and questioned by researchers. The fundamental principle of the index-based method is that a project's past performance is the best available indicator of future performance (Kim and Reinschimdt 2010). However, cost-forecasting with these conventional formulas is typically unreliable early in a project because of the few data points available (Fleming and Koppelman 2006; Zwikael et al. 2000). These two limitations show premise for further extensions of EVM-related cost-forecasting.

Earned-Value Management Cost Forecasting with a Regression-Based Method

As an alternative to the index-based approach, various techniques based on regression analysis have been gaining recognition as valuable methods to support the cost-forecasting activity. In regression analysis, a dependent variable (a response, typically the AC) is regressed against an independent variable (a predictor, typically time) to compute the CEAC. The regression model can be either linear or nonlinear to represent the respective relationship between the response and predictor (Nystrom 1995). Regression-based studies overcome the limitations inherent with index-based techniques and thus are available for wider boundaries of application. The parameters of a regression model found through a regression analysis represent the behavior of a project with respect to the entire lifecycle (Alvarado 2004). Moreover, even though the regression-based computation effort is greater compared with the relatively simple index-based cost-forecasting method, it yields better estimates early in the project life (Christensen and Heise 1993; Tracy 2005).

Growth Models

Regression-based techniques are grounded on the notion and theories associated with the growth models. Growth models have been frequently used to study population growth in fields such as biology, economics, and marketing. These models describe situations inherent to data with a growth pattern, in which the growth rate monotonically increases to a maximum before it steadily declines to zero (Seber and Wild 1989). This behavior

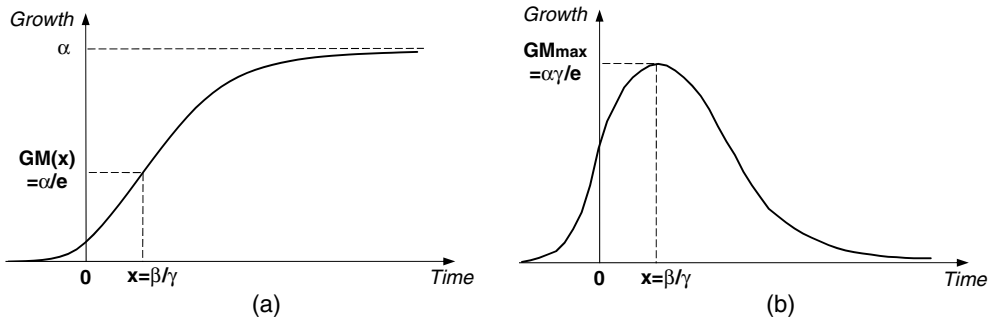


Fig. 2. Characteristics of an s-shaped growth model: (a) cumulative-growth curve; (b) growth rate curve

is well-described by an s-shaped or sigmoidal pattern that is extensively used in curve-fitting and forecasting of population growth. Such models are characterized for the position of the point of inflection being the time at which the growth rate is the greatest. Fig. 2 shows the typical characteristics of growth models.

In the research reported in this paper, the writers apply four types of the three-parameter growth model, as follows: (1) logistic model (LM), (2) Gompertz model (GM), (3) Bass model (BM), and (4) Weibull model (WM). Table 1 introduces their generic cumulative distribution functions (CDF) that are used for curve-fitting and forecasting. Table 1 also provides the parameterized CDF equations and specific mathematical properties set for the research reported in this paper.

All of these growth models can be used to describe the cost expenditure behavior of construction projects because their functional form and parameters reflect the nature of physical improvement progress and satisfy the requirements for a typical s-shaped cost pattern of construction projects. During the initial stages of a project, construction progress is typically slow-paced because of field preparation, equipment deployment, and excavation works. By midlife the construction progress speeds up, increasing the work rate to a maximum, and finally decreases, declining the work rate to zero during the completion phase.

A common mathematical feature of these models is that they all have an α value to represent the asymptotic project final cost as time (x) approaches infinity. In other words, as a construction project tends to its completion, there is less work left to accomplish and the finishing phase is a typical slow-paced approach to the final cost (the α-asymptote).

The differences in mathematical properties and behaviors render these models applicable to a variety of fields, as presented in a diverse and large body of literature (Bates and Watts 1988; Hines and Montgomery 1990; Seber and Wild 1989). As part of this vast amount of literature, the writers next present a brief description and review of previous applications of the four previously noted growth models to the specific issue of forecasting in construction-project management.

The LM is one of the most widely used s-shaped growth models because of its simplicity and analytic tractability (Seber and Wild 1989). The LM has the same parameters and meanings of the GM. However, the LM is normally distributed, having the inflection point at 50% of total growth, whereas the GM has it at approximately 1/3 of the total growth. De Marco et al. (2009) applied the logistic curve to forecast the TEAC of a construction project. Their model extrapolates the nonlinear s-shaped curve of EV data of an ongoing project and allows the evaluation of time estimates by considering the specific initial behavior of a project based on current schedule progress. Their project case proved that the TEAC computed by using the LM is more accurate than the estimate computed by using the index-based formula.

The GM is often used to study population growth in cases for which the growth curve is not symmetrical and typically right-skewed (Seber and Wild 1989). Inflection happens at time $x = \beta/\gamma$ with cumulated growth $\mathrm{GM}(x) = \alpha/\mathrm{e}$ (approximately 1/3 of the final cost) when the cost expenditure rate reaches its maximum $G_{\max} = \alpha\gamma/e$ (Fig. 2).

In both the LM and GM, a β parameter represents the y-intercept initial size and γ is a scale parameter that governs the rate of growth. Trahan (2009) developed GM curves to forecast a project's CEAC by using EVM data of a number of U.S. Air Force acquisition contracts from 1960 to 2007. Trahan (2009) used the GM parameters (found by regressing the values of the cumulative AC up to a project completion against corresponding values of time points) to determine a remainder of BAC at 20% completion and then to predict a project's CEAC. The results of Trahan (2009) are similar to previous studies (Christensen et al. 1995) in that the

Table 1. Growth Models and Mathematical Properties

Property	Logistic	Gompertz	Bass	Weibull
Generic CDF	$\alpha/[1+\beta e(-kx)]$ (Seber and Wild 1989)	$\alpha e\{-e[-k(x-\gamma)]\}$ (Seber and Wild 1989)	$\{1-e[-(p+q)x]\}/\{1+(q/p)e[-(p+q)x]\}$ (BBRI 2010)	$1-e\{-[(t-\gamma)/\delta]^{\beta}\}$ (Hines and Montgomery 1990)
Parameterized CDF	$\mathrm{LM}(x)=\alpha/[1+e(\beta-\gamma x)]$	$\mathrm{GM}(x)=\alpha e[-e(\beta-\gamma x)]$	$\mathrm{BM}(x)=\alpha(\{1-e[-(\beta+\gamma)x]\}/\{1+(\gamma/\beta)e[-(\beta+\gamma)x]\})$	$\mathrm{WM}(x)=\alpha\{1-e[-(x/\gamma)^{\delta}]\}$
Inflection point	$x=\beta/\gamma$; $\mathrm{LM}(x)=\alpha/2$	$x=\beta/\gamma$; $GM(x)=\alpha/e$	$x=\ln[(\gamma/\beta)/(\beta+\gamma)]$; $\mathrm{BM}(x)=\alpha[0.5-(\frac{\beta}{2\gamma})]$	$x=(1/\gamma)(\delta-1/\delta)^{1/\delta}$; $\mathrm{WM}(x)=1-e[-(1-\delta^{-1})]$
Symmetry	Symmetrical	Asymmetrical	Asymmetrical	Flexible
Maximum growth rate	$\alpha\gamma/4$	$\alpha\gamma/e$	$\{\alpha[(\beta+\gamma)^2]\}/4\gamma$	$1/\gamma$
Parameters	α = asymptote β = y-intercept γ = scale	α = asymptote β = y-intercept γ = scale	α = asymptote β = y-intercept γ = scale	α = asymptote γ = scale δ = shape

growth models may not be more accurate than index-based methods to compute the CEAC in all cases. However, the model appeared to be appropriate and accurate to forecast the final cost of either overrun or overrun-close projects.

The BM (Bass 1969) has been used in marketing science for over 4 decades. The model explains the process of how new products are adopted and the rate at which they penetrate the marketplace. It provides good forecasts of the total sales, its peak, and timing of the peak based on historical data (Satoh et al. 2001). Application studies of this model (Table 1 shows a generic CDF) in the construction industry by Alvarado et al. (2004) showed that the BM best works to develop the EV baseline curves based on its curve-fitting procedure. This is because the behavior of the cumulative EV follows the s-shaped pattern in which it is cumulated at a constant p (units of cost) per time point and the growth rate increases by q (units of cost) per time point for each p earned. Based on historical projects, BM is a reliable technique to model the EV baseline curves for future construction, repair, and alteration projects (Alvarado et al. 2004).

The WM is a widely-used model in the field of manufacturing and reliability engineering, failure analysis, and weather forecasts (Hines and Montgomery 1990). Unlike the other three growth models, the WM is extremely flexible because of its δ-shaped parameter. By changing the value of the parameter, the WM distribution can be related to a number of other distributions. For instance, when the parameter value is 1.0 the model represents the exponential distribution, and when its value is 3.4 it approximates the distribution that is normal, representing the LM curve. Its scale parameter represents the growth rate and governs the inflection point of the cost expenditure, whereas the shape parameter determines its distribution. The WM has wide application in cost control and analysis within U.S. defense projects. Brown (2002), based on 128 research and development projects, found that curve-fitting with WM well-describes a project budget profile and developed Weibull-based regression models to forecast the required δ-shaped and γ-scale parameters. Nassar et al. (2005) applied Weibull analysis to evaluate reliability of a project's schedule performance utilizing CPI and SPI data. They concluded that the advantage of the WM relative to the conventional index-based approach is that the Weibull analysis provides accurate performance analysis and risk predictions with a relatively small number of data points, rendering it to best work as early as three data points available.

Earned-Schedule Method

The methodology proposed in this paper integrates these given growth models with ES. Developed by Lipke (2003), ES is regarded as one of the primary methods for schedule analysis and time forecasts (PMI 2011). The advantages of ES compared with EV are that it allows measuring the schedule indicators in time instead of currency units and eliminating the deficiency of the SPI tending to 1.0 as the project approaches completion even though it is behind or ahead of schedule. Hence, the ES concept introduces a new time-based SPI, also denoted as SPI(t). The ES is found by comparing EV against PV and its value is calculated by plotting the EV curve at actual time (AT) to the PV curve (Fig. 1). This projection point represents the time in which the current EV should actually have been earned. Depending on whether a project is ahead or behind its schedule, this time point can be after or before AT, respectively. The following set of equations [Eqs. (3)–(7)] are used in this paper to analyze a project's progress and provide time forecasts:

$$\mathrm{ES} = C + I \tag{3}$$

where C represents the number of time increments for which EV exceeds PV; and I represents the following:

$$I = \frac{\mathrm{EV} - \mathrm{PV}_c}{\mathrm{PV}_{c+1} - \mathrm{PV}_c} \tag{4}$$

$$\mathrm{SPI}_{(t)} = \frac{\mathrm{ES}}{\mathrm{AT}} \tag{5}$$

$$\mathrm{TEAC} = \frac{\mathrm{PD}}{\mathrm{SPI}_{(t)}} \tag{6}$$

$$\mathrm{CF} = \frac{\mathrm{TEAC}}{\mathrm{PD}} = \frac{1}{\mathrm{SPI}_{(t)}} \tag{7}$$

where CF = completion factor, which indicates forecasted time-completion yielded to unity.

When the value of the CF is 1.0, based on current work progress, this indicates that the duration of a project is as planned; more than 1.0, a project is likely to be finished with schedule delay; and less than 1.0, an early finish. Comparative studies by Henderson (2004), Henderson and Zwikael (2008), and Vandevoorde and Vanhoucke (2006) proved relative accuracy in forecasting the TEAC by the ES method. With regard to construction projects, De Marco and Narbaev (2013) proved that the TEAC found by using ES is more accurate than the EV-based formula and provides a better early estimate of the total duration of a project.

Methodology

The CEAC methodology proposed in this paper develops a new formula to forecast the final cost of an ongoing construction project, effective for early and middle stages of its execution. The formula interpolates the parameters of the growth models found through nonlinear regression analysis, and then it is refined by integrating a value of the ES-based CF to consider any schedule progress and time estimates that might affect the construction-project cost performance.

To achieve this purpose the methodology is accomplished in three steps. First, the writers combine PV and AC data versus time to develop the growth models and determine their parameters. Because all growth models have an intrinsic nature of a nonlinear relationship between a predictor (input) and response (output), this step is performed through nonlinear regression analysis. Second, the writers interpolate the results of the growth models into the CEAC equation. Third, considerations of the final time completion and schedule progress are taken into account and hence this formula is refined by integrating a value of the ES-based CF. Fig. 3 presents the three-stepped modeling process of the proposed CEAC methodology together with the tasks to be performed in each step.

Step 1: Developing the Growth Model

The first step is to develop the regression-based nonlinear growth model that was used to fit the cumulative cost s-shaped curve line of a construction project. For demonstration purposes, the LM is used and its equation parameterized, as per Eq. (8)

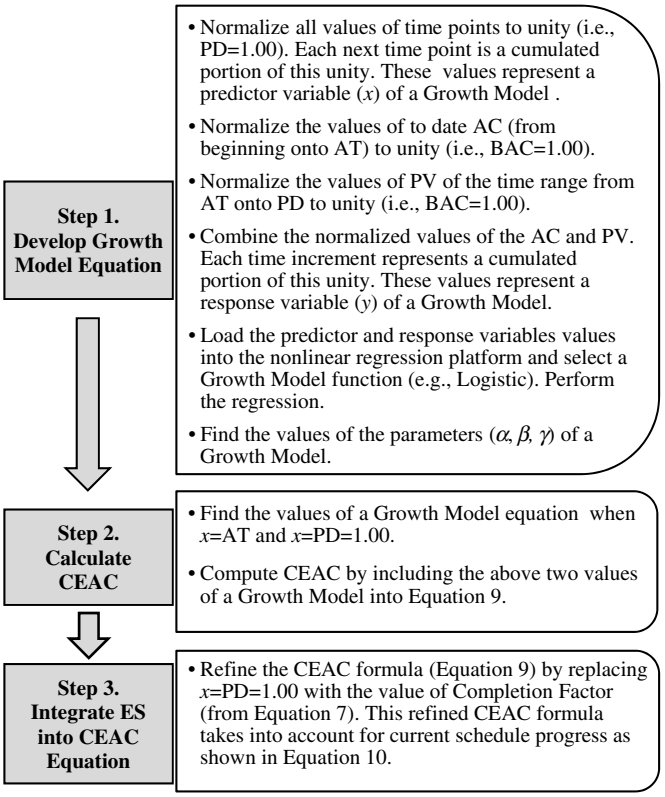


Fig. 3. Cost estimate at completion methodology modeling-process

$$\mathrm{LM}(x) = \frac{\alpha}{1 + e(\beta - \gamma x)} \tag{8}$$

Eq. (8) has three parameters, as follows: (1) the α-asymptote represents the never-attained value of the final cost as time tends to infinity, (2) β is the y-intercept (which represents the initial size of a project cumulative cost), and (3) the γ-scale parameter that governs the rate of the cost growth. The maximum cost growth rate is $G_{\max} = a\gamma/4$ and it occurs when the time $x = \beta/\gamma$ and the cumulative cost $\mathrm{LM}(x) = \alpha/2$. In Eq. (8) the predictor x represents normalized time points with its maximum value equal to 1.00 (100% time-complete); i.e., the PD of a project. The corresponding value of the response variable is normalized points of the combined-to-date AC (from a project beginning to AT) and PV (from AT and onto 100%) with its maximum value (BAC) of 1.00. This allows building the combined dataset of AC and PV with respect to time points used in the model-fitting and determining the three parameters for the LM nonlinear equation.

The last task in this step is to define an equation for the writers' LM through nonlinear regression analysis. For this purpose, best-fit values of model parameters were obtained to fit the model data and then compared with the other three models (Bates and Watts 1988). The Minitab software package is used. Unlike linear regression, building a nonlinear regression model requires specifying the starting values for a model parameters and the algorithm for approximation of these values (Bates and Watts 1988). However, no standard procedure exists to define good starting values. Unless one knows the starting values of the model parameters based on prior information, to define a good set of these values in nonlinear model fitting is difficult because it is because of the existence of a nonlinear relationship between a predictor and a response variable and the phenomena the model represents. To estimate values of these parameters, both linear and nonlinear regression use the least squares (LS) method of approximation. The most common assumption in curve-fitting is that data points are randomly scattered around an ideal curve with the scatter in accordance with a Gaussian distribution (Bates and Watts 1988). Taking into account this consideration, the LS approach minimizes the sum of the squared errors (SSE, the difference between the estimated values and actual input values of the parameters) of the vertical distances of the points from a curve. The research reported in this paper applies a Gauss-Newton algorithm for the LS approximation, which obtains convergence iteratively close-to-linear regression that are not heavily dependent on the starting values. The iteration process continues until the algorithm converges to determine the parameter values within the specified tolerance on the minimum SSE (Bates and Watts 1988). In the research reported in this paper, the candidate model is regarded as statistically valid if the Newton-Gauss algorithm of the LS approximation converges to estimate values of the growth model's parameters within specified convergence tolerance. When the minimization algorithm fails to perform this task, the model is rejected from further CEAC comparative study.

In the writers' regression analysis, the levels of the confidence interval (CI) are set, i.e., lower and upper CI, which are lower and upper 100% (1–0.05) endpoints of the estimates, respectively. The CI gives a range of estimate values between two limits in which the actual values are more likely to fall. The CI represents the accuracy of an estimate and the 95% confidence level is regarded as an accepted standard (Seber and Wild 1989).

Step 2: Calculating the Project CEAC

To compute the CEAC the writers used Eq. (9), which assumes the values of the growth model when a project is to-date and 100% complete. This formula is similar to the classical formula in Eq. (1) because both equations have AC. However, in Eq. (9) the remaining portion of the CEAC is computed based on the nonlinear growth model results, whereas Eq. (1) corrects the remaining portion of the BAC by CPI. Trahan (2009) presents the generic form of Eq. (9) and developed the nonlinear growth model by regressing the response values of the AC for the entire project lifecycle against the corresponding time increments. Unlike the approach of Trahan (2009), the research reported in this paper combines values of current AC and PV, as presented previously in step 1

$$\mathrm{CEAC}(x) = \mathrm{AC}(x) + [\text{growth model}(1.0) - \text{growth model}(x)](\mathrm{BAC}) \tag{9}$$

Knowing the three parameters for the equation defined previously in step 1, the writers computed the value for the growth model 1.0 in which the time is 100% complete and it trends to its α-asymptote. For the growth model x, a predictor variable is the to-date value of the time increment. These two values are calculated by using Eq. (8) and their difference represents the estimated portion of BAC needed to complete a project, which is in turn added to the current-to-date AC incurred to compute the final CEAC.

Step 3: Integrating ES into a CEAC Equation

This step considers a project final cost affected by schedule progress; hence, Eq. (9) is refined to take this assumption into account. This is achieved by replacing the value of the time $x = 1.00$ for the growth model function by a value of CF (the ratio of the to-date TEAC to a project PD) found using the ES method.

This correction to the CEAC formula is introduced as given in Eq. (10)

$$\text{CEAC}(x) = \text{AC}(x) + \{\text{growth model}[\text{CF}(x)] - \text{growth model}(x)\}(\text{BAC}) \quad (10)$$

Validation

Applicability

The proposed CEAC methodology is demonstrated through application on nine past projects to construct various types of civil, industrial, infrastructure, and residential facilities internationally. These projects all have medium-sized budgets with average BAC close to US$8 million and planned completion times varying from 6 to 27 months (the average PD is 13.3 months). These case projects and their EV records are reported in the literature. Table 2 provides the list of these projects along with their associated information and EV data. Five out of nine projects experienced cost overruns and seven projects reported a schedule delay. In most cases the number of time points and corresponding AC points used for nonlinear modeling is either four or five (including the initial time point $x = 0$); this shows that the nonlinear regression is computed based on a small sample size. Regression analysis is useful early in a project life when little or unreliable EV performance and progress data are available.

The advantage of regression modeling in this context is that it extrapolates past available data with future planned data, whereas the conventional EV approach solely relies on past performance and progress. This extrapolation is achieved through the development of the growth model in which the values of its parameters show a relationship between past, current, and future project performance and progress. The CF shows the forecasted progress outcome and based on its value it suggests that seven out of nine projects finished with schedule delay.

The writers next demonstrate the three-step procedure on a numerical example, project A, an industrial facility renovation project, and then show results for all cases. Step 1 of the methodology is about developing the regression-based nonlinear growth model that will be used to fit its s-shaped curve to the cumulative cost curve of project A. Table 3 shows both the initial input data and the results of step 1. After 3 months of execution, the project is 20.76% complete and therefore this is the period in which the CEAC is computed. Eq. (11), presented in the next paragraph, shows the nonlinear LM equation for project A.

Minitab developed Eq. (11) based on the options the writers set for the nonlinear regression analysis, as discussed next. The writers define good starting values for the three parameters. Taking into account the normalization to unity of both the predictor and response variables the writers define 1.0 as a starting value for all parameters. The confidence level is then set at 95% with the Gauss-Newton algorithm to converge on the minimum SSE. The maximum number of iterations is 200 with a default-set convergence tolerance of 1×10^{-5}. Fig. 4(a) presents the LM-fitted s-curve of project A. The curve fits the AC-PV data of the project very well; i.e., all eight response values are in the CI (upper and lower dashed curves). The writers are confident with a probability of 95% that the CI estimates the real AC-PV data

$$\text{LM}(x) = \frac{1.134}{1 + e(4.463 - 6.302x)} \quad (11)$$

Eq. (11), the LM equation, fits an output of percent complete (response of the fitted curve) with an input of time complete (predictor of the fitted curve). The model suggests that at month 3, the cumulative cost is 0.222 of the BAC and this happens when the project is 48.4% time-complete. The asymptote of the LM

Table 2. Case Projects and EVM Data

Project	Type	Duration (months)		Outcome	AT (months)	CF
		Planned	Actual			
A (De Marco et al. 2009)	Renovation	6.2	7	CO-LF	4	1.133
B (Shokri-Ghasabeh and Akrami 2009)	Infrastructure	15	16	CO-LF	5	1.083
C (Khamidi et al. 2011)	Infrastructure	27	31	CO-LF	4[a]	1.022
D (Vandevoorde and Vanhoucke 2006)	Renovation	9	13	CU-LF	4	1.093
E (Ibid)	Renovation	9	12	CO-LF	4	1.134
F (Ibid)	Renovation	10	9	CO-EF	4	0.982
G (Valle and Soares 2006)	Civil	10	10	OB-OS	5	1.154
H (Singletary 2006)	Civil	13	14	CU-LF	5	1.846
I (Cheng et al. 2010)	Residential	21	24	CU-LF	10	0.996

Note: CO = cost overrun; CU = cost underrun; EF = early finish; LF = late finish; OB = on-budget; and OS = on schedule.
[a]Quarterly increments.

Table 3. Step 1 Results for Project A

Time points	Months	Predictor time	EV complete (%)	Real AC-PV values		Fitted AC-PV values	
				Cumulative	Growth rate	Cumulative	Growth rate
1	0	0.000	0.00	0.000	0.000	0.013	0.013
2	1	0.161	0.58	0.025	0.025	0.035	0.022
3	2	0.323	6.23	0.115	0.090	0.092	0.057
4	3	0.484	20.76	0.213	0.098	0.222	0.130
5	4	0.645	33.65	0.450	0.237	0.465	0.243
6	5	0.806	52.57	0.750	0.300	0.737	0.272
7	6	0.968	69.01	0.920	0.170	0.949	0.212
8	6.2	1.000	100.00	1.000	0.080	0.978	0.029

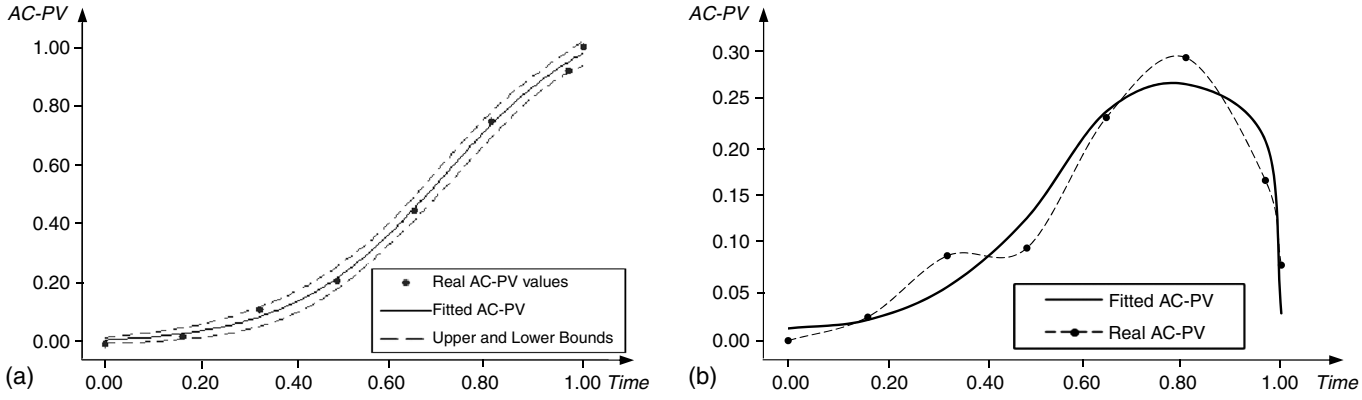


Fig. 4. Logistic model s-shaped curve of project A: (a) fitted curve with a confidence interval of 95%; (b) growth rate curve

equation shows that as the project time approaches infinity the project has a 13.4% cost overrun. From the mathematical properties of Eq. (11), the LM equation (Table 1), the inflection point occurs when the project is 70.8% time-complete and the cost is 56.7% of the BAC. The cost-growth rate of the project starts with 0.013% of the BAC for the first 1% of time-complete and it increases until the inflection point, at which it reaches its maximum of 1.787% per 1% time-complete. Table 3 shows monthly adjusted growth rates. The cost-expenditure rate of the fitted curve line is at its maximum 5 months into the project, when it is 27.2% of the BAC, whereas the PV is 30.0% for that month [Fig. 4(b)].

When the project is from 48.40 to 100% time-complete, the nonlinear fitting procedure allows computing the cumulative actual cost as 0.222 and 0.978 of the BAC, respectively (Table 3). Furthermore, in accordance with step 2 the writers calculated the CEAC using Eq. (9). The computed CEAC is 5.30% less than the final actual cost and project A was delivered with a cost overrun of 2.34%.

The final step requires integration of the value of the CF to consider the effects of the schedule progress into the project's cost. The CF for project A is 1.133 (Table 2). In Eq. (9) the value of the time $x = 1.00$ is replaced by this value of the CF [Eq. (10)]. The 2.78% value of the new ES-based CEAC is slightly more than the final actual cost. However, this refined estimate with integration of the CF is more accurate than the underestimated value obtained without considering the CF as per Eq. (9), i.e., –5.30%.

Comparison of the Growth Models and an Index-Based Method

The third objective of this paper is to determine which growth model among the four studied works best to forecast CEAC in both early and middle stages of project execution. Two criteria are set to perform this task, as follows: (1) statistical validity of the growth model equation, and (2) accuracy of the CEAC. Therefore, the writers first evaluated if the LS algorithm finds a solution for the parameter values of the growth model equation through the Minitab regression platform and selected the growth model for which the parameter values were found with respect to all nine cases. A comparative study of their CEACs is then provided. The accuracy of the estimates of the equations is based on (1) a comparison of percentage error (PE), which is termed the difference between the actual and estimated values of final cost expressed as a percentage; and on (2) the mean absolute percentage error (MAPE) of the number of valid projects tested. These two measures were computed in accordance with Eqs. (12) and (13)

$$\text{PE}\% = \frac{\text{CEAC} - \text{CAC}}{\text{CAC}} \times 100 \tag{12}$$

$$\text{MAPE}\% = \frac{\sum |\text{PE}|}{n} \tag{13}$$

where CAC = cost at completion; and n = number of statistically valid projects.

For early-stage estimates, Table 4 presents both PE and MAPE of the CEAC computed using the four models and it shows the original source's forecast computed with the index-based method. Four cases fail to converge. The same analysis is conducted for middle-stage estimates (45–65% complete) with eight cases failing in convergence. Minitab determines the cause of this failure as multicollinearity in the model. This problem occurs in multiple

Table 4. Estimate Errors and MAPE of the CEAC in the Early Stage, 10–30% Complete

	LM		GM		BM		WM		
Project	Base	ES	Base	ES	Base	ES	Base	ES	CPI-based
A	−5.30	2.78	−11.08	−0.19	−5.55	3.44	FtC	—	−0.29
B	−7.76	−5.17	−6.96	−2.91	−6.55	−0.07	−6.48	−0.16	16.26
C	−6.78	−1.66	−5.74	−4.37	−6.42	2.87	−5.74	6.53	−1.01
D	1.32	2.67	1.55	5.21	1.28	2.91	0.98	2.53	−9.31
E	−14.85	−12.71	−14.57	−9.67	−15.02	−11.31	−14.89	−9.75	−4.11
F	−6.79	−6.95	−5.18	−5.59	−6.81	−7.01	−7.00	−7.16	2.02
G	−7.63	−6.33	−4.03	−0.48	−7.31	−5.89	−8.01	−7.52	−28.51
H	−9.58	−7.48	−6.68	0.54	−9.38	−6.03	−8.20	−5.40	−1.58
I	FtC	—	9.72	8.74	FtC	—	FtC	—	12.24
MAPE	7.50	5.72	7.27	4.20	7.29	4.94	7.32	5.57	8.37

Note: FtC indicates that the model failed to obtain the convergence within a tolerance of 1×10^{-5}.

regressions when two or more predictors are highly linearly correlated, which causes erratic changes in the parameter estimates (Bates and Watts 1988). Several measures are available to solve this issue, such as simplifying the model equation, transforming predictors, parameterization, and so on (Seber and Wild 1989). However, this issue is beyond the scope of this paper; the failing growth model equation is considered invalid and is discarded in the further CEAC comparative study.

The writers next performed a comparison of the accuracy of the final cost estimates. For early-stage estimates, in the base case, i.e., without applying the ES method, the results of the estimates error show that the GM has the smallest MAPE of 7.27 (Table 4). However, it is difficult to recognize its accuracy because of negligible differences in the MAPE of the four models. Comparison of the individual PE results yields nothing in the base cases. The application of the ES concept into the CEAC methodology allows clearing some points. First, the results of individual PE show that the GM outperforms other models in five cases out of eight (the writers eliminated project I because it is the sole model in which the GM works). Second, the comparison of MAPE results shows that the GM is improved with a value of 4.20%. However, individual PE suggests that the GM does not produce explicitly more accurate results than the other three models. The CEAC calculated by the CPI-based formula has largest MAPE of 8.37% and its PEs are more unsteady compared with regression-generated CEACs. Table 5 provides final MAPE for the middle-stage estimates of the nine projects in which again, at the control point, the GM ES-based equation provides the smallest result (3.44%) compared with the CPI-based and GM-based methods. Overall, the results of this comparative study allow concluding that the GM is a statistically valid model that works for all nine projects in both early stages and midstages, and it provides more accurate CEAC when its equation is integrated with ES.

In addition, the writers address a study on the impact of the GM's and the CPI-based method's factor on CEAC at three consecutive time points (i.e., $x-1$, x, and $x+1$). The reason for performing this kind of sensitivity analysis is to see if changes in CF and CPI with respect to time have a respective influence on CEAC accuracy for the GM ES-based and index-based methods. The results confirm this statement. First, both factors (CF and CPI), with changing values with respect to time, have a positive influence on the CEAC accuracy. Second, the sensitivity test proves that the GM provides more steady errors in the estimates close to actual outcomes, whereas the CPI-based method's CEACs have larger differences. This large difference in errors is associated to CPI instability in the traditional approach, which is solely dependent on past cost performance

The research reported in this paper validates the GM using a set of industry projects and compares it with the traditional CPI-based method. In total, the writers screened 34 construction projects

Table 5. Change in MAPE at Three Consecutive Time Points

	Model input		Model output		
Time point	CF	CPI	GM base	GM ES-based	CPI-based
Early, average	1.21	0.99	6.53	4.85	10.72
Precontrol	1.36	0.99	6.59	6.52	16.27
Control	1.16	1.02	7.28	4.20	8.37
Postcontrol	1.12	0.98	5.72	3.83	7.52
Middle, average	1.12	0.99	5.01	4.55	4.85
Precontrol	1.11	0.96	5.14	4.14	4.41
Control	1.10	0.98	5.11	3.44	4.87
Postcontrol	1.14	1.03	4.79	6.08	5.28

Table 6. MAPE of the CEAC for Industry Projects

Execution stage	Projects	GM base	GM ES-based	CPI-based
Early, 10–30%	21	6.10	4.93	8.17
Middle, 45–65%	26	5.63	3.21	6.38

carried out from 1998 to 2010 by an Italian construction company that calls for anonymity. The projects that fail meeting EVM application requirements (lack of PV, EV, and/or AC; incomplete or partial reporting; and lack of schedule progress data) were excluded from the test. Twenty-one projects were selected for early-stage and 26 for middle-stage forecasts. Table 6 presents the final MAPE results in which the GM outperforms the CPI-based method in calculating CEAC. In particular, in the early stage, the CPI-based method produces a MAPE of 8.17, whereas the GM base case (−6.10%) with an even more accurate result of 4.93% when it considers schedule progress.

In the middle-stage estimate, the result evidences the same trend. However, the estimate errors for all three models are closer to actual cost outcome than in early-stage forecasts. This is because of greater uncertainty during a project initial stage getting reduced as it progresses to completion with the values of the CF and CPI tending to stabilize.

Implications, Limitations, and Future Research

The research reported in this paper originates some considerations inherent with both theoretical and practical implications. It paves the way to the integration of index-based methods with regression-based CEAC models that so far have been considered as two separate streams of project-management research. In particular, the methodology combines both actual and future planned values, and provides for accurate and reliable early estimates. In this sense, the proposed approach helps to overcome the intrinsic limitations of index-based methods, such as backward-looking on past EV information and unreliable cost-forecasting early in a project life. As a result, the theory developed in this paper is a contribution to the evolution of EVM body-of-knowledge through a combination of statistical analysis.

The research reported in this paper has some practical implications. First, the proposed methodology is a cost-schedule integrated approach, which provides a viable and accurate CEAC-forecasting method because the cost estimate considers the schedule impact as a determinant factor of cost behavior. Second, it may also be used for small-sized and short-duration construction project because the nonlinear regression analysis works as early as three time points available in a short-lifespan project.

However, like all EVM systems, the proposed method has a prerequisite inherent with the availability of progress, cost, and schedule-measurement information reported by the project team at all levels of the organization. Unless reliable and timely reporting is established, the proposed CEAC methodology, as any other EVM method, will only provide false cost-forecasts (Fleming and Koppelman 2002, 2004).

The proposed methodology gives rise to a number of interesting issues for future research. First, the theoretical background of the developed methodology calls to provide comparative analyses between the proposed model and conventional index-based methods with respect to stability and timeliness of cost-forecasting.

Second, the paper limits further analysis of the schedule progress and duration forecasts bounding the involvement of the ES method with contribution of the CF only. Therefore, a third area of future research is associated with the application of other

concepts of schedule factors, such as impact of schedule changes, schedule fast-tracking, sequential relationships, and dynamic scheduling.

The project-management community uses other techniques that produce accurate results for both cost and duration forecasts. Hence, the promising avenue for future research is on comparative study of the GM with such techniques as artificial intelligence-based tools [artificial neural network (ANN), fuzzy logic models, and so on], Bayesian beta s-shaped curve, Kalman filter, and other valuable and promising methodologies.

To ease diffusion and usage of the developed methodology, a computer tool must be developed. Accordingly, the next practice-oriented objective is to create application software that may assist construction industry practitioners in the desired implementation of the writers' technique.

Conclusion

This paper proposes a new nonlinear regression-based model to assist project managers in the task of forecasting the final cost at completion of a construction project that is early into its life. The method enhances forecasting capabilities of the EVM technique by combining conventional index-based forecasting approaches and statistical regression analysis. It is a cost-schedule integrated approach that interpolates characteristics of an s-shaped growth model with the ES technique to calculate the CEAC of a construction project. This issue is regarded as a research novelty that contributes to the extension of the EVM body-of-knowledge.

The new methodology was developed to provide more accurate and reliable CEAC forecasts effective for early and middle stages of project execution. Three objectives are accomplished to achieve this purpose, as follows: (1) a theoretical basis for a new CEAC formula was formulated; for this, the writers grounded into the research context both index-based and regression-based cost-forecasting methods, the four growth models, and the ES concept; (2) to prove viability and practicability of the formula it was validated on nine construction projects; and (3) the writers selected the best-performing growth model by testing their statistical validity and comparing the accuracy of their estimates. The validation and comparison revealed that the GM is a statistically valid model and generates more accurate CEAC estimates than those computed by using the other three models and the index-based method.

References

Alvarado, C. M., Silverman, R. P., and Wilson, D. S. (2004). "Assessing the performance of construction projects: Implementing earned value management at the General Services Administration." *J. Facil. Manage.*, 3(1), 92–105.

Anbari, F. T. (2003). "Earned value project management method and extensions." *Proj. Manage. J.*, 34(4), 12–23.

Bates, D. M., and Watts, D. G. (1988). *Nonlinear regression analysis and its applications*, Wiley, New York.

Bass, F. (1969). "A new product growth for model consumer durables." *Manage. Sci.*, 15(5), 215–227.

Bass' Basement Research Institute (BBRI). (2010). "Mathematical derivation of the Bass model." ⟨http://www.bassbasement.org/BassModel⟩ (Nov. 19, 2012).

Brown, T. W. (2002). "Forecasting research and development program budgets using the Weibull model." M.S. thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH.

Cheng, M. Y., Peng, H. S., Wu, Y. W., and Chen, T.-L. (2010). "Estimate at completion for construction projects using evolutionary support vector machine inference model." *Automat. Construct.*, 19(5), 619–629.

Christensen, D. S., Antolini, R. D., and McKinney, J. W. (1995). "A review of estimate at completion research." *Cost estimating and analysis*, Springer, New York, 207–224.

Christensen, D. S., and Heise, S. (1993). "Cost performance index stability." *Natl. Contract Manage. J.*, 25(1), 7–15.

De Marco, A., Briccarello, D., and Rafele, C. (2009). "Cost and schedule monitoring of industrial building projects: Case study." *J. Construct. Eng. Manage.*, 10.1061/(ASCE)CO.1943-7862.0000055, 853–862.

De Marco, A., and Narbaev, T. (2013). "Earned value-based performance monitoring of facility construction projects." *J. Facil. Manage.*, 11(1), 69–80.

Fleming, Q. W., and Koppelman, J. M. (2002). "Earned value management: Mitigating the risks associated with construction projects." *Program Manage.*, 31(2), 90–95.

Fleming, Q. W., and Koppelman, J. M. (2004). "Sarbanes-Oxley: Does compliance require earned value management on projects." *Contract Manage.*, 33(4), 26–28.

Fleming, Q. W., and Koppelman, J. M. (2006). *Earned value project management*, 3rd Ed., Project Management Institute, Newtown Square, PA.

Henderson, K. (2004). "Further developments in earned schedule." ⟨http://www.earnedschedule.com/Docs/Further%20Developments%20in%20Earned%20Schedule.pdf⟩ (Aug. 22, 2013).

Henderson, K., and Zwikael, O. (2008). "Does project performance stability exist? A re-examination of CPI and evaluation of SPI(t) stability." ⟨http://www.earnedschedule.com/docs/does%20project%20performance%20stability%20exist%20%20henderson%20zwikael.pdf⟩ (Aug. 22, 2013).

Hines, W. W., and Montgomery, D. C. (1990). *Probability and statistics in engineering and management science*, 3rd Ed., Wiley, New York.

Khamidi, M. F., Ali, W., and Idrus, A. (2011). "Application of earned value management system on an infrastructure project: A Malaysian case study." *Proc., Int. Conf. Manage. Service Sci.*, International Association of Computer Science and Information Technology, Singapore, 1–5.

Kim, B. C., and Reinschmidt, K. F. (2010). "Probabilistic forecasting of project duration using Kalman filter and earned value method." *J. Construct. Eng. Manage.*, 10.1061/(ASCE)CO.1943-7862.0000192, 834–843.

Lipke, W. (2003). "Schedule is different." ⟨http://earnedschedule.com/Docs/Schedule%20is%20Different.pdf⟩ (Aug. 22, 2013).

Lipke, W., Zwikael, O., Henderson, K., and Anbari, F. (2009). "Prediction of project outcome: The application of statistical methods to earned value management and earned schedule performance indexes." *Int. J. Proj. Manage.*, 27(4), 400–407.

Minitab version 16 [Computer software]. Minitab, State College, PA.

Nassar, K. M., Gunnarsson, H. G., and Hegab, Y. M. (2005). "Using Weibull analysis for evaluation of cost and schedule performance." *J. Construct. Eng. Manage.*, 10.1061/(ASCE)0733-9364(2005)131:12(1257), 1257–1262.

Nystrom, T. D. (1995). "A comparative study of linear and nonlinear estimate at completion methods." M.S. thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH.

Project Management Institute (PMI). (2008). *A guide to the project management body of knowledge*, 4th Ed., Newtown Square, PA.

Project Management Institute (PMI). (2011). *Practice standard for earned value management*, 2nd Ed., Newtown Square, PA.

Satoh, D. (2001). "A discrete bass model and its parameter estimation." *J. Oper. Res. Soc. Jpn.*, 44(1), 1–18.

Seber, G. A. F., and Wild, C. J. (1989). *Nonlinear regression*, Wiley, New York.

Shokri-Ghasabeh, M., and Akrami, N. (2009). "How does a new set of earned value management schedule control work? A case study in Iran." *Trans. Environ. Dev.*, 5(2), 136–145.

Singletary, M. R. (2006). "Assessing the financial feasibility of implementing wireless technologies for construction management." M.S. thesis, Florida State Univ., Tallahassee, FL.

Tracy, S. P. (2005). "Estimate at completion: A regression approach to earned value." M.S. thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH.
Trahan, E. N. (2009). "An evaluation of growth models as predictive tools for estimates at completion (EAC)." M.S. thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH.
Valle, J. A., and Soares, C. A. P. (2006). "The use of earned value analysis (EVA) in the cost management of construction projects." *Proc., Project Management Institute Global Congress*, Newtown Square, PA, 1–11.
Vandevoorde, S., and Vanhoucke, M. (2006). "A comparison of different project duration forecasting methods using earned value metrics." *Int. J. Proj. Manage.*, 24(4), 289–302.
Zwikael, O., Globerson, S., and Raz, T. (2000). "Evaluation of models for forecasting the final cost of a project." *Proj. Manage. J.*, 31(1), 53–57.