A New Method to Improve Quality Predicting of Software Project Completion Level

The-Anh Le[1,2], Quyet-Thang Huynh[1(✉)], and Thanh-Hung Nguyen[1]

[1] School of Information and Communication Technology, Hanoi University of Science and Technology, Hanoi, Vietnam
{thanghq,hungnt}@soict.hust.edu.vn
[2] Faculty of Information Technology, People's Police University, Bacninh, Vietnam

Abstract. Earned Value Management (EVM) is a powerful tool for estimating costs and evaluating a software project. Many methods have been used to improve the effectiveness of EVM in evaluating a software project, in which the method of applying the Gompertz growth model (GGM) is one of the effective directions. The paper studies the method of combining the Gompertz growth model and the earned value management method (GGM-EVM) to predict the cost to complete the software project. The team experimented with modeling a number of software project data sets in practice, running and testing and analyzing the results. Several improvements have been proposed to increase the effectiveness of the GGM-EVM method, resulting in relatively positive research results.

Keywords: Project management · EVM · Growth models · Gompertz

1 Introduction

Earned value management (EVM) is one of the well-known techniques for controlling the time and cost of a project [1, 6]. This method is based on a set of metrics to measure and evaluate the overall health of a project in order to provide an early warning to the project administrator of project problems. However, this method has some limitations such as: based only on past costs, the prediction lacks reliability in the early stage of the project and does not take into account forecasting statistics [6]. These three limitations are the main reason for the development of new methods [1, 6]. One of the methods is the use of linear or nonlinear regression analyzes to develop regression models, also known as GM-Growth Models [2].

Currently, there are many models that predict the cost of completion of a project, as well as models that predict the end of a project. Different models have been adequately studied and compared with Batselier, J et al. In [4]. Some related studies using EVM method can be mentioned as: Khamooshi and Golafshani (2014) in [5] proposed new method EDM improved than most ESM methods, while Lipke and Watt (2011) in [6] and Elshaer (2013) in [7] extended the old ESM method which was very effective in the early phase of the project, but less effective in the later stages of the project. The authors

N.-S. Vo et al. (Eds.): INISCOM 2021, LNICST 379, pp. 211–219, 2021.
https://doi.org/10.1007/978-3-030-77424-0_17

Narbaev T.; De Marco A. (2014) [3] proposed a method that combines growth models and EVM methods for some positive results.

There are a number of studies related to project completion time forecasting and project completion cost prediction by applying different performance factors in the combined growth model and EVM approach to improve forecast quality [4, 6, 8]. Each of the models above has its own advantages and disadvantages and is applicable to specific datasets. In the framework of this paper, we focus on research on growth model Gompertz and apply on the data of real projects in [9]. We inherit the results of the methods that have been studied in the article [2, 4, 8] by testing and evaluating the experimental results, then propose to improve the value management technique to obtain the results with Gompertz growth model to improve the quality of project completion prediction. [10] proposed a method for improvement the parameter estimation of non-linear regression in growth model to predict project cost at completion.

The next content in the paper is presented as follows: Sect. 2 presents the value management method and growth model Gompertz; Sect. 3 presents project completion cost estimation and proposes I-GGM methodology to improve predictive quality; Sect. 4 presents project completion time prediction and suggested improved I-Regression method; Finally, in Sect. 5 presents conclusions, scientific contributions and development directions of the next research.

2 Background

2.1 The Earned Value Management Method

UML is most generally used to provide a standard way to visualize a system's design and is also widely used for test case generations. There are many types of research in recent years about various techniques for generating test cases from UML diagrams.

Earned Value Management is an efficient tool used to predict project completion time and cost based on current project status. EVM has 06 main parameters as follows [6]:

- PV (Planned Value): Value as planned, PV = BAC *% of expected work.
- AC (Actual Cost): Actual cost, which is the actual cost spent at the time of monitoring the project.
- EV (Earned Value): The earned value, EV = BAC *% of the actual work, is the sum of the PV values that have been completed, from the start of the project to the time of project monitoring.
- ES (Earned Schedule): Time as planned, is the time spent by AC according to PV plan.
- BAC (Budget at Completion): Funding to complete the project.
- PD (Plan Duration): Project duration.

Project performance in terms of time and cost, determined by comparing the key parameters PV, AC, EV, and ES provides the following performance measurement results:

– CPI (Cost Performance Index): Cost performance index. Formula for calculating CPI = EV / AC;
– SPI (Schedule Performance Index): Index of planned performance. The formula for calculating SPI = EV / PV;
– SPI (t): Adjustment plan performance index (symbol (t) to specify that this formula relates to time). Formula to calculate SPI (t) = ES / AT;

Cost Estimate at Completion (CEAC) is calculated by the following formula:

$$CEAC = AC + PCWR = AC + \frac{(BAC - EV)}{PF} \tag{1}$$

Where:
AC: Actual cost at the present moment (i.e. actual time AT).
PCWR: Estimated costs for the remaining works (estimates for the future).
PF (Performance Factor): Performance factor.

The CEAC estimation methods based on the different efficiency coefficients of the EVM model are shown in Table 1.

Table 1. Overview of EVM-based project completion cost prediction methods

	CEAC1	CEAC2	CEAC3	CEAC4	CEAC5
SPI	PF = 1	PF = CPI	PF = SPI	PF = SCI	PF = 0.8 * CPI + 0.2 * SPI
SPI(t)			PF = SPI(t)	PF = SCI(t)	PF = 0.8 * CPI + 0.2 * SPI(t)

Estimated time to complete the project (TEAC - Time Estimate at Completion) is calculated [4.6]:

$$\text{TEAC} = AT + PDWR \tag{2}$$

Where:

AT: Current time;
PDWR: Estimated time of remaining jobs, calculation is also based on PE coefficient.

To predict the time to complete the project, one of three methods based on PV, ED and ES [4] can be used.

2.2 Gompertz Growth Model (GGM)

Gompertz function is used in combination with EVM to increase efficiency estimates TEAC and CEAC, this function is often used to describe the phenomenon with model S-shaped growth [2, 4]:

$$G(t) = \alpha \exp[-e^{\beta - \gamma t}] \tag{3}$$

where α, which represents the asymptotic value ($t \rightarrow \infty$) of the Gompertz function and is therefore related to the final budget of the project. The parameter γ, which characterizes the growth rate of the cumulative curve, allows the study of a variety of different project cost profiles

For $\beta = \gamma T$, where T is the vertex of the distribution function, we have the Gompertz function and the distribution function:

$$G(t) = \alpha \exp[-e^{-\gamma(t-T)}] \tag{4}$$

$$g(t) = \frac{dG(t)}{dt} \alpha\gamma \mathrm{G}(t) e^{-\gamma(t-T)} \tag{5}$$

We can define the end of the project as a specific part of the asymptotic value, such as 95% or 99%, in that case, at the intended end of the project.

$$G(T_1) = (1-\varepsilon)G(t \rightarrow \infty) = (1-\varepsilon)\alpha = \alpha exp[-e^{\beta-\gamma t}] \tag{6}$$

where ε is a constant. We define k satisfying:

$$1 - \varepsilon = exp[-e^{-k}] \tag{7}$$

deduce:

$$k = \lambda(T_1 - T) \tag{8}$$

Therefore, k is determined when we choose a specific endpoint of the project, eg, 99% of the asymptotic, α.

3 Research Methodology

3.1 Original Method

The authors in [3] have proposed the ES-GGM method to predict the cost of completing the project, including 3 steps as follows:

Step 1: Construct an S-curve of the growth model

The first step is to build an S-curve that combines actual cost (AC) and expected value (PV).

With the curve function GGM as follows:

$$G(t) = \alpha \exp[-e^{\beta\gamma - T)}] \tag{9}$$

Use the least square regression algorithm to find three parameters of the GGM model: (α, β, γ) such that the smallest deviation from the reference data. With CEAC prediction, we need to construct a GGM curve regression for the value AC - the actual cost to predict AC at the end. Input data because if only taking AC until the time of making prediction, it will be too little and easily cause error in prediction. Therefore, in [1, 2], the authors take the PV data portion of the remaining time to compensate for the missing data. Specifically, the reference data as input to the regression are as follows:

AC data until the time of making prediction t.

PV data from the time of the prediction to the end of the project.

This data set is inserted as input to the least square regression algorithm and the output is a trio of parameters of the GGM model: (α, β, γ).

Step 2: Calculate project completion cost based on Gompertz growth model (GGM)

Performing CEAC calculation according to GGM:

$$CEAC(x) = AC(x) + [GGM(1.0) - GGM(x)] * BAC \tag{10}$$

Step 3: Calculate the cost of completing the project based on GGM plus ES:

According to ES-GGM:

$$CEAC(x) = AC(x) + [GGM(1/SPI(t)) - GGM(x)] * BAC \tag{11}$$

3.2 Proposed Method I-GGM

Because the calculation process of the regression model uses many PV values, the generated curve is very close to these values. This leads to when predicting according to GGM, the results are usually quite close to PV, which is mostly inaccurate.

We propose a method to improve the I-GGM. In which, instead of regression according to AC + PV, we will regress according to AC and the rest is the AC value predicted using EVM-CPI. Thus, GGM will calculate the trend of data via CPI and help improve the accuracy of the solution.

The improvement method I-GGM proposed by us will include the following steps:

Step 1: Construct an S curve of the growth model

The first step is to construct an S-curve that combines the actual cost (AC) and the AC value predicted through the CPI.

The S curve is constructed as a combination of the actual cost curve (AC) and the remainder the estimated AC values based on the CPI. This curve will reflect the past and future of the project.

In order to construct an S-curve and apply a growth model, it is necessary first to standardize project data as follows:

- Standardize all project time point values in terms of units (ie PD will be normalized to 1.00). Each successive timeline will be the cumulative part of this unit. These normalized values will be the predictor (x) variable of a GGM growth model.
- Standardize the actual cost values AC (from the beginning of the project to the present time AT) in terms of units (ie BAC = 1.00).
- Standardize the expected values of AC according to the CPI (from the time of AT to the PD of the project) in terms of units (BAC = 1.00).

After normalization, we proceed to combine the normalized values of actual AC and expected value AC above. Then we get a time-axis S curve from the inception to the

PD of a project, which is a combination of AC (historical data) and expected value AC according to CPI (future data) of project data.

Step 2: Calculate the growth model parameters Gompertz: (α, β, γ).

Step 3: Calculate the cost of completing the project: Calculate the CEAC according to the formula (12)

$$CEAC(x) = AC(x) + [GGM\,(1/SPI(t)) - GGM\,(x)] * BAC \tag{12}$$

4 Experiment and Evaluation

4.1 Experimental Data

The actual project datasets are listed from the website: https://www.projectmanagement.ugent.be/research/data/realdata. In which, we have performed data filtering statistics to select 20 projects in the field of software engineering. The experimental data are shown in Table 2.

Table 2. Data on 20 projects in the field of software engineering

ID	Project name	BAC (€)	Duration (month)	tracking
1	C2011-07 Patient Transport System.xlsx	180759.44	12.9666	23
2	C15-10 Tax Return System (1).xlsx	18990	2.8333	3
3	C2015-11 Staff Authorization System.xlsx	14400	1.8333	3
4	C2015-12 Premium Payment System.xlsx	132570	6.1333	3
5	C2015-13 Broker Account Conversion System.xlsx	12735	3.9	4
6	C2015-14 Supplementary Pensions Database.xlsx	34260	4.1347	4
7	C2015-15 FACTA System.xlsx	11700	1.9	3
8	C2015-16 Generic Document Output System.xlsx	64620	9	6
9	C2015-17 Insurance Bundling System.xlsx	281430	6.9388	5
10	C2015-18 Tax Number system (2).xlsx	39450	4.2666	3
11	C2015-20 Policy Numbering System.xlsx	12645	5.7	4
12	C2015-21 Investment Product (1).xlsx	4020	1.2333	2
13	C2015-22 Risk Profile Questionnaire.xlsx	29880	5.0361	4
14	C2015-23 Investment Product (2).xlsx	46920	4.0388	4
15	C2015-24 CRM System.xlsx	44130	7.7666	6

(*continued*)

Table 2. (*continued*)

ID	Project name	BAC (€)	Duration (month)	tracking
16	C2015-25 Beer Tasting.xlsx	1210	0.4402	3
17	C2015-26 Debt Collection System.xlsx	458112.3683	4.9388	5
18	C2015-28 Website Tennis Vlaanderen.xlsx	219275	6.7	4
19	C2016-08 SCM System.xlsx	375253.343	24.1666	34
20	C2016-09 Data Loss Prevention System.xlsx	584951.7688	6.5	9

4.2 Evaluation Criteria

a) **Percentage error measure (PE)**
The PE reflects the efficiency of each method in predicting the cost of completing the project. The PE is the difference between the actual cost and the estimated cost as a percentage:

$$PE\% = \frac{(CEAC - AC)}{AC} 100\% \tag{13}$$

where:

CEAC: estimated cost at the time of project completion.
AC: the actual cost of completing the project.

b) **Mean Absolute Percentage Error measure (MAPE)**

The Mean Absolute Percentage Error (MAPE): corresponding to the average value of PE on different projects is calculated by the formula:

$$MAPE\% = \frac{100\%}{n} \sum\nolimits_{i=1}^{n} \frac{|CEAC_i - AC_i|}{AC_i} = \frac{1}{n} \sum\nolimits_{i=1}^{n} |PE_i|\% \tag{14}$$

4.3 Result of Assessment According to MAPE

The calculated MAPE value is evaluated through 5 main phases:

- The beginning of project, Tracking 0;
- The first half of project,Tracking ¼;
- Between Project Tracking ½;
- The second half of project, Tracking ¾;
- At the end of project, the Tracking last.

At each phase, average the PE value across all 20 projects with each algorithm.

The MAPE results when predicting the CEAC of the 5-phase methods from 0 to 4 are shown in Fig. 1. The horizontal axis represents the project phases and the vertical axis shows the value of MAPE (%).

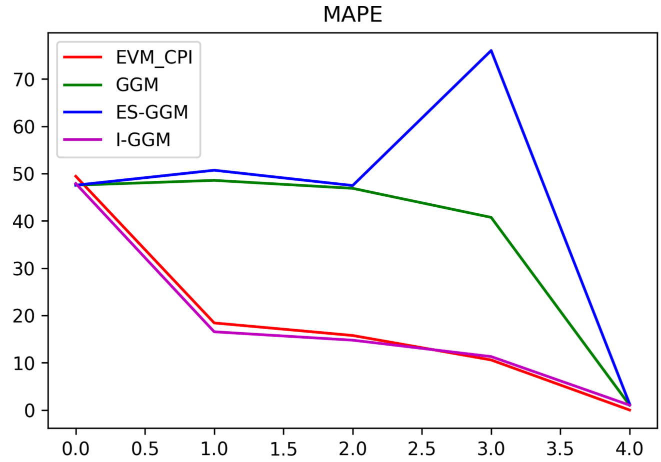

Fig. 1. MAPE results when predicting CEAC

In which, EVM-CPI predicts EVM-CPI, GGM and ES-GGM are predicted by GGM regression using the method in [1, 2], and I-GGM is the proposed improved regression model. Realizing that, GGM and ES-GGM showed ineffective results compared to EVM, while the improved I-GGM model was effective compared to EVM. The explanation given is similar to the above, partly because the CPI volatility of software engineering projects is too high, making it harder to predict GGM meth-od. Besides, the fact that projects have a relatively low tracking number also significantly affects the regression quality. After applying the innovation, I-GGM better captured the uptrend by CPI to give more stable results than using pure GGM and ES-GGM. Considering the average MAPE results obtained in the above figure, I-GGM can be rated slightly better than EVM-CPI (due to the higher and more significant predictive importance at the beginning of the project. than predicted at the end of the project).

5 Conclusions

In this paper, a methodology was proposed that combines Gompertz growth model and EVM techniques to improve the quality of predicting software project completion. The improvement method I-GGM uses a cost-performance index (CPI) that predicts future costs instead of the PV value that builds the S curve closest to reality to predict the cost of completion most accurate project. Conducted experiments on 20 data sets of actual software projects published on the website [9]. The test results show significant and relatively good improvements of the I-GGM improvement methods compared to the previous models such as GMM, ES-GGM, EVM. Overall, the results obtained confirm that the proposed improvement has theoretical basis and is justified by experiment with actual data sets.

Development direction: In the future, we will continue to improve the algorithm and collect more real data for more accurate testing.

Acknowledgments. This research is funded by Hanoi University of Science and Technology under Grant number T2018-TĐ-009.

References

1. Simion, C.P., Marin, I.: Project cost estimate at completion: earmed value management versus earned schedule-based regression models. A comparative analysis of the models application in the construction projects in Romania. Econ. Comput. Econ. Cybern. Stud. Res. **52**(3) (2018)
2. Nannini, G., Warburton, R.D.H., De Marco, A.: Improving the accuracy of project estimates at completion using the Gompertz function. In: International Research Network on Organizing by Projects (IRNOP) 2017, UTS ePRESS, Sydney: NSW, pp. 1–15 (2017)
3. Narbaev, T., De Marco A.: Combination of growth model and earned schedule to forecast project cost at completion. J. Constr. Eng. Manag. **140**(1), Article number 04013038 (2014). ISSN 0733–9364.
4. Batselier, J., Vanhoucke, M.: Evaluation of deterministic state-of-the-art forecasting approaches for project duration based on earning value management. Int. J. Proj. Manag. **33**(7), 1588–1596 (2015)
5. Khamooshi, H., Golafshani, H.: EDM: earned duration management, a new approach to schedule performance management and measurement. Int. J. Proj. Manag. **32**, 1019–1041 (2014)
6. Fleming, Q.W., MeKoppelman, J.: Earned value project management. Project Management Institute Newtown Square, Pennsylvania, USA (2015)
7. Elshaer, R.: Impact of sensitivity information on the prediction of project's duration using earning schedule method. Int. J. Proj. Manag. **31**(4), 579–588 (2013)
8. Le, T.A., Huynh, Q.T., Nguyen, T.H., Nguyen, N.H., Cao, P.N.: Correction for future performance factor PF in EVM-GM method evaluating the completion of software projects: testing and evaluation. In: Proceedings of the Association National Workshop XXI: Selected issues of Information and Communication Technology, pp. 137–143. Scientific and Technical Publishing House, Hanoi (2018)
9. Batselier, J., Vanhoucke, M.: https://www.projectmanagement.ugent.be/research/data/realdata
10. Huynh, Q.T., Le, T.A., Nguyen, T.H., Nguyen, N.H., Nguyen, D.H.: A method for improvement the parameter estimation of non-linear regression in growth model to predict project cost at completion. In: 2020 RIVF International Conference on Computing and Communication Technologies. IEEE (2020)