

Examining Project Duration Forecasting Reliability

Walt Lipke
PMI® Oklahoma City Chapter

Abstract

Earned Schedule (ES) forecasting of project duration has been researched over several years. Overwhelmingly, in comparison to other EVM-based methods, ES has been affirmed to be better. However, the testing results from a study, which employed simulation techniques, indicated there were conditions in which ES performed poorly. These results create skepticism as to the reliability of ES forecasting. This paper examines that study, focusing on the unfavorable results. The analysis put forth indicates ES forecasting to be more reliable than portrayed by the study and is enhanced by the application of the longest path method.

Introduction

A research study of project duration forecasting was made several years ago, employing simulation methods applied to created schedules having several variable characteristics (Vanhoucke & Vandevorde, 2007). Three Earned Value Management (EVM) based methods were compared in the study: 1) EVM¹, Earned Duration (ED)², and 3) Earned Schedule (ES)³. The overall result from the study was that forecasts using ES, on average, are better than the others. However, in certain instances the ES forecast was not. This result appears counter-intuitive due to the fact that, by its formulation, ES forecasts must converge to the actual final duration. Because of this apparent discrepancy, the conditions of the study are examined for an explanation.

For the study the researchers developed a full range of possible schedule performance scenarios against which the simulations were examined. These scenarios are depicted in figure 1. However, in the research publication, the scenarios were incompletely described. There was insufficient description of several key components of the scenario model. What is meant by the symbols (-, 0, +) in the ovals at the top and left side of the diagram? Likewise, there is lack of definition of the terminology, “Critical activities,” and Non-critical activities.”

It is believed the researchers’ intent of the figure 1 diagram was to show, in a very succinct way, combinations of performance factors (symbols and type of activity) with their associated outcomes, SPI(t)⁴ and project duration. However, the lack of clarity in the research paper in describing the

¹ EVM duration forecasting is accomplished by dividing the planned duration (PD) by the Schedule Performance Index (SPI). Reference (Project Management Institute, 2011).

² Reference (Jacob & Kane, 2004)

³ Reference (Lipke, 2009)

⁴ “SPI(t)” is the time-based schedule performance index. For the research study it is a term used to represent, in general, the index for all three forecasting methods examined.

meaning of the symbols and activity type has clouded the understanding of the results and conclusions drawn.

Model Description

The scenario model as shown in figure 1 has nine possibilities. The possibilities are determined from the pairing of the symbols (-, 0, +) between the critical and non-critical activities. For example, “-” for critical activities can be paired with “-,” “0,” and “+” for non-critical activities. Thus, with three pairings for each critical activity symbol, we understand why there are nine scenarios.

What are these symbols? What do they represent? ...The symbols are briefly described in the paper to indicate a general condition of schedule performance:

- “-” better than expected
- “0” as expected
- “+” poorer than expected

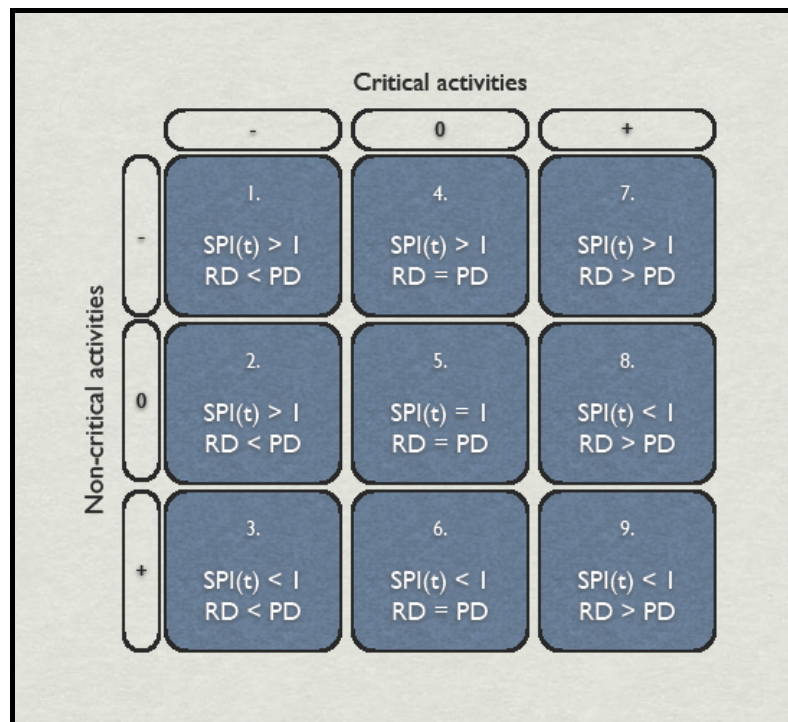


Figure 1. Schedule Performance Scenarios⁵

These performance characteristics are the components of the nine possible pairings. Nevertheless, it is unclear as to why these pairings are necessary in studying the capability of the three previously cited forecasting methods. It would seem all that is required is to use the output of the simulations

⁵ Reference (Vanhoucke M. , 2008)

for the analysis. Apparently, however, the researchers must have believed additional comparison information could be derived from the various performance scenarios.

To achieve the pairings the researchers forced the performance conditions into the simulations. Possibly the tactic can be rationalized, but it appears contrary to how a simulation is normally employed. Usually, simulation is used because it is just too difficult to analyze the system directly. In this instance, the researchers have perturbed the system and consequently the results. This raises the question, “Do these results accurately portray the forecasting performance of the various methods, when the performance conditions are contrived?”

To this point, “Why is it necessary to analyze forecasting performance by segregating critical and non-critical activities?” Additionally, these components of the study model have not been clearly defined. That is, “What is the meaning of the terms “critical” and “non-critical” activities?” It is believed that the term “critical” infers the tasks or activities that are on the critical path within the schedule.⁶ Thus, “non-critical” connotes those tasks that are not on the critical path.

This understanding of the two terms is used for the remainder of this paper. Nevertheless, there is a degree of ambiguity in that the critical path can change during project execution. It is presumed the researcher used “critical” in reference to the critical path of the planned schedule, which in the absence of a schedule change is invariant with execution. This assumption is made because of the difficulty the researchers would have inducing the performance conditions into the simulation while simultaneously accounting for the changes in the critical path.

Performance Scenario Analysis

The performance results from the various scenarios are partitioned in the research paper into three categories:

- True
 - Misleading
 - False
- The true scenarios (1, 2, 5, 8, 9)⁷ have the characteristic that the relationship of the real or final project duration (RD) to the planned duration (PD) can be inferred from the schedule performance efficiency indicator, SPI(t). Using scenario 1 for example, SPI(t) is greater than 1 (indicating good performance), while RD is less than PD (as one would expect from the indicator); i.e., the indicator is consistent with the duration result.
- The misleading scenarios (4, 6) are characterized by the critical activities being completed as planned, while the non-critical activities are not. The RD equals PD; however, SPI(t) is either greater or less than 1. Thus, the indicator is inconsistent with the duration outcome.

⁶ Critical path, as usually defined, is the longest path of the schedule.

⁷ The numbers in parenthesis refer to the nine numbered cells of figure 1.

- The false scenarios (3, 7) occur for two circumstances: 1) When non-critical activity performance is good and critical performance is poor, or 2) When critical activity performance is good and non-critical is poor. For these scenarios, the indicator, SPI(t), infers an outcome in opposition to the actual duration.

Figure 2 is the compiled results from the simulations for each of the nine scenarios. As shown, ES forecasting is vastly superior to the other methods for the scenarios in the true category. However, ES performs poorly for both the misleading and false categories. These findings raise the question, “Although ES is the preferred forecasting method, as concluded by the research, “How can a project manager know the forecast is reliable?” The first impression the various scenarios provide is that ES forecasting is extremely unreliable. Of the nine possible outcome scenarios, four have poor correlation between SPI(t) and the actual project duration.

In wrestling with the response, a more general question is considered initially, “Can project managers reasonably expect forecasts to be one-hundred percent consistent with final outcomes?” Certainly forecasting has some expectation of error; it is a prediction of the future, which has a goodly amount of uncertainty. Thus, project managers have some expectation that their tools are imperfect; therefore some error is accepted. Nevertheless, as a profession, we strive for perfection and work to improve from the flaws discovered in our methods. Accordingly, through examination of the anomalous performance reported in the study, effort is focused on improving ES forecasting and building confidence in its application.

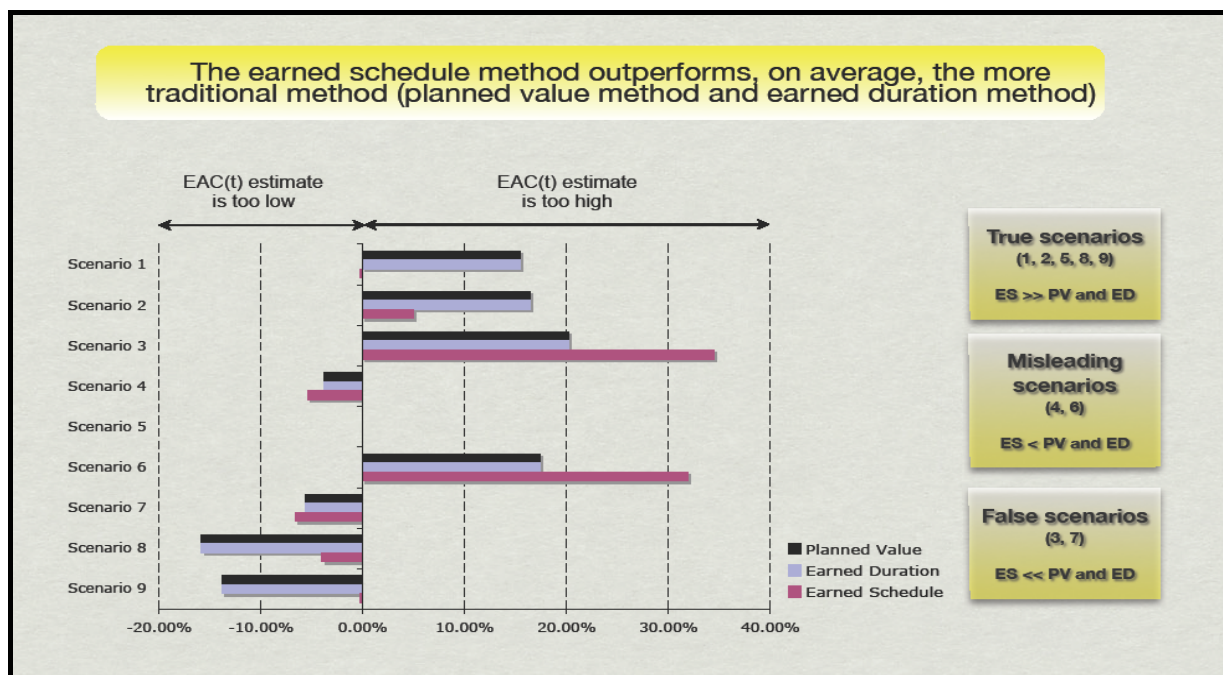


Figure 2. Scenario Performance Results⁸

The concern for the four scenarios, in which ES forecasting does not perform well, leads to an examination of the performance conditions. In reviewing figure 1 some observations are made:

⁸ Reference (Vanhoucke M. , 2008)

- RD always matches critical activity performance
- SPI(t) correlates with non-critical activity performance

Let us review these connections in more detail. For critical activities:

- “-” correlates to $RD < PD$
- “0” correlates to $RD = PD$
- “+” correlates to $RD > PD$

And, for non-critical activities:

- “-” correlates to $SPI(t) > 1$
- “+” correlates to $SPI(t) < 1$

By recognizing and examining these correlations, it may be possible to gain insight as to the reasons for the ES forecast performance inconsistencies. As mentioned earlier, it is quizzical that ES does not perform well, when its forecast always converges to the actual project duration.

It is curious as to why there are observed correlations between the symbols, RD, and SPI(t). Because the observations are consistent, it is hypothesized that there may be additional conditions or constraints imposed in the study. A condition/constraint which would explain the connection between the symbols and RD for critical activities is to confine the project to complete on the critical path of the planned schedule. As a matter of opinion, this is an unrealistic condition and thus raises questions as to the validity of the study results. Projects do not always complete on the planned critical path.

The second connection correlating the “+” and “-” symbols to SPI(t) implies that the amount of planned/earned value used in computing the indicator is greater for the non-critical activities than for the critical activities. This can be deduced from the fact that SPI(t) represents the project performance, as a whole; SPI(t) values align with the non-critical symbols only when the volume of work for the non-critical activities is predominant. In general, this volume of work relationship between critical and non-critical activities is likely a true condition; but, it may not be as the schedule topology becomes increasingly serial. Applying this distribution relationship of planned/earned value to the critical and non-critical activities is seen as unnecessary and most likely perturbs the simulation results.

Affirming the Study Finding

Assuming the preceding analysis explaining the correlations is accurate, thereby raising concerns with the study, “Can the research paper’s conclusion that ES forecasting is the best method be upheld?” Certainly, the paper is supported by other studies and application using real data (Lipke, 2008) (Crumrine & Ritschel, 2013). That should be sufficient; however, it is desirable to use the discovery from the scenario analysis with the study results to draw the same conclusion.

Of course, the study could be re-performed without the imposed conditions and constraints. The result would most likely support the paper’s conclusion; even so, we would like to avoid expending effort performing additional simulations and analysis. One possibility in minimizing effort is to review the misleading and false scenarios to see if, in some way, they over emphasize ES unreliability. If this is the case, then the remaining true scenarios are predominant, thereby increasing emphasis on the study conclusion of ES being the best EVM-based forecasting method.

A logical argument is to recognize that performance of the non-critical activities affects the performance of the critical activities. For instance, a non-critical activity may need to be completed before a critical activity can begin. When the non-critical activity lags in its performance, the dependent critical activity will, in all likelihood, lag, as well. The misleading and false scenarios imply little connection between the performance of non-critical and critical activities. This is indicated by the correlation between critical activity performance and RD, and the total lack of correlation with non-critical performance. This unrealistic condition can be imposed by forcing, as the researchers did, but cannot be sustained throughout the execution of a real project due to task inter-dependency. As well, for ES forecasting the misleading and false indication conditions will resolve to SPI(t) and RD agreement during execution because of the characteristic of convergence to the final duration.

To further amplify the deductions from the foregoing analysis, the evolution of scenario categories is illustrated in figure 3, using notional data. The figure demonstrates the influence of task inter-dependency and the convergence quality of ES forecasting. As is readily seen from the “True” graph, ES forecasting becomes increasingly reliable as the project proceeds to completion. As “True” is increasing, the components contributing to unreliability, “Misleading” and “False,” are decreasing.

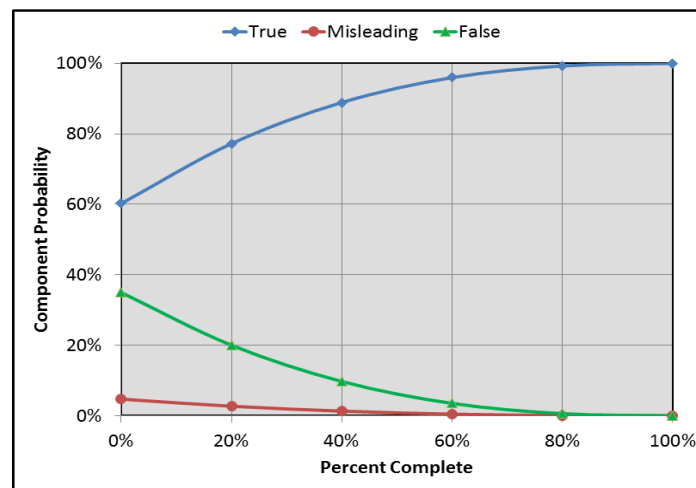


Figure 3. ES Forecasting Reliability

Of course the graph for all components (True, Misleading, False) and their relative probabilities will differ with changes to the topological structure of the schedule; forecasts for serial schedules are the most reliable, while parallel are least. Thus, the figure is representative, only. However, the graphs are credible in that they were created conservatively by beginning with True at 60 percent, approximating its percentage (55%) of total scenarios.

Although figure 3 is notional, the interpretation of increasing forecasting reliability it illustrates bolsters confidence in the ES method. For example, using the True graph, the probability of a reliable forecast when the project is 20 percent complete is 77 percent. Later in the performance, when the project is 80 percent complete, the forecast made at that time is 99 percent reliable. In relation to the beginning point of 60 percent reliability, ES forecasting is demonstrated to improve rapidly with project accomplishment.

From the preceding analysis the findings from the 2007 study have been affirmed and positive argument is made for ES forecasting. Furthermore, ES is logically assessed to be more reliable than the perception created from the four misleading and false scenarios presented by the study.

Extending the analysis, an observation is made that primarily serial schedules have much less opportunity for misleading and false conditions to occur. For a completely serial schedule, SPI(t) must describe performance on the critical path, and hence inconsistency between the performance indicator and RD is significantly diminished. This observation explains the results from the simulation study indicating ES forecasting is more reliable for serial topology schedules.

The discussion thus far has been concerned with ES forecasting for the total project. Recently, an advancement in ES theory hypothesized significant forecasting improvement by applying the methodology to the “longest path”⁹ (Lipke, 2012). ES forecasts using LP, termed “ES-LP,” are virtually unaffected by the topology of the network schedule. This advancement reduces the possibility of misleading and false forecasting, thereby increasing ES reliability. With ES-LP, it is impossible to have the condition of poor, or as expected, critical performance while non-critical is good, yielding a disconnect between SPI(t) and RD. While SPI(t) for the total project indicates good overall performance, SPI(t) for the longest path cannot when critical path performance is poor. Thus, for longest path, the instances of SPI(t) in disagreement with RD are significantly reduced for scenarios 4 and 7.

Additionally, application of longest path forecasting reduces the disagreement for scenarios 3 and 6. It is a near impossibility to have the disconnect between SPI(t) and RD for these scenarios. Regardless of planned critical path performance (good or as expected), poor schedule performance for longest path indicates there is a longer path to completion. Thus, in reference to scenarios 3 and 6, performance on the planned critical path is irrelevant in forecasting the final duration. However, it remains possible, though more difficult, to have SPI(t) values for the longest path which do not coincide with RD.

Therefore it should be clear, the enhancement of ES forecasting offered by the use of the longest path method causes SPI(t) to be significantly more consistent with outcome duration. Longest path forecasts are more focused into the true scenarios. ES-LP forecasting, theoretically, is the most reliable EVM-based method.

Summary and Conclusion

The 2007 forecasting study contains curious results. Questions have lingered for some time as to why ES forecasting performed exceedingly well for the majority of scenarios in the study, but for

⁹ The “longest path” (LP) is the serial path within the network schedule having the longest duration forecast, using ES methods.

some the method was very poor. Although the study concludes that ES forecasting is better than other EVM-based methods, it provides a negative view of ES reliability.

The analysis of the induced conditions discussed the causes for the misleading and false scenarios of the study. The lack of task inter-dependency required to produce these scenarios is shown to be unrealistic. Subsequently, it is argued that the reality of inter-dependency between critical and non-critical activities will not allow performance on the planned critical path, solely, to dictate the final duration.

As illustrated by figure 3, the misleading and false scenarios are inherently unstable. The inconsistency between SPI(t) and RD from the misleading and false scenarios is shown to be overcome by the evolution of project performance to the true scenarios. Therefore, the study conclusion is upheld by this examination: *ES provides better forecasts than other EVM-based methods*. Furthermore, it is established that ES reliability increases as the project moves toward completion. Consequently, ES forecasting is more reliable than perceived from the study.

Finally, the application of longest path is hypothesized to minimize the misleading and false scenarios, improving both accuracy and reliability of ES forecasting.

Final Comments

The conclusions in this paper are logically derived and would welcome confirmation. Experimenters are challenged to perform simulations without the scenario control imposed in the 2007 study. Following is a description of the suggested experiments:

Using the results from the uncontrolled simulations, tabulate the natural distribution of scenario occurrences. The scenarios are depicted in figure 4 as nine scenarios of project duration outcome versus the ES performance indicator, SPI(t). These scenarios no longer have the distinction of critical and non-critical activities. As shown, the True scenarios are 1, 5, and 9; Misleading scenarios are 2, 4, 6, and 8, while False scenarios are 3 and 7. Group the results to their appropriate scenario and to the True, Misleading, and False categories. Perform this procedure at 10 percent increments of project completion. As well, repeat the above directions for schedules of varying serial/parallel topology. For comparison, perform the operations and analysis for both ES and ES-LP forecasting.

The scenario and group tabulations are expected to illustrate the dominance of the True scenarios in a more natural environment, thereby confirming the conclusion that project duration forecasting reliability is greater than perceived from the study. Furthermore, these experiments are projected to conclude that ES and, especially, ES-LP duration forecasting are very reliable methods.

| | | Outcome | | |
|-----------|------------|--------------|--------------|--------------|
| | | RD < PD | RD = PD | RD > PD |
| Indicator | SPI(t) > 1 | 1 True | 4 Mislead | 7 False |
| | SPI(t) = 1 | 2 Mislead | 5 True | 8 Mislead |
| | SPI(t) < 1 | 3 False | 6 Mislead | 9 True |

Figure 4. Indicator vs Outcome Scenarios

References:

Crumrine, K., & Ritschel, J. (2013). A Comparison of Earned Value Management as Schedule Predictors on DOD ACAT 1 Programs. *The Measurable News*, Issue 2: 37-44.

Jacob, D., & Kane, M. (2004). Forecasting schedule completion using earned value metrics ...revisited. *The Measurable News*, (Summer) 1, 11-17.

Lipke, W. (2008). Project Duration Forecasting: Comparing Earned Value Management Methods to Earned Schedule. *CrossTalk*, (December) 10-15.

Lipke, W. (2009). *Earned Schedule*. Raleigh, NC: Lulu Publishing.

Lipke, W. (2012). Speculations on Project Duration Forecasting. *The Measurable News*, Issue 3: 1, 4-7.

Project Management Institute. (2011). *Practice Standard for Earned Value Management, 2nd Edition*. Newtown Square, PA: PMI.

Vanhoucke, M. (2008). Measuring Time: a simulation study of earned value metrics to forecast total project duration. *Earned Value Analysis Conference 13*. London.

Vanhoucke, M., & Vandevoorde, S. (2007). A simulation and evaluation of earned value metrics to forecast the project duration. *Journal of the Operational Reserach Society*, Issue 10, Vol 58: 1361-1374.

About the Author



Walt Lipke

Oklahoma, USA



Walt Lipke retired in 2005 as deputy chief of the Software Division at Tinker Air Force Base. He has over 35 years of experience in the development, maintenance, and management of software for automated testing of avionics. During his tenure, the division achieved several software process improvement milestones, including the coveted SEI/IEEE award for Software Process Achievement. Mr. Lipke has published several articles and presented at conferences, internationally, on the benefits of software process improvement and the application of earned value management and statistical methods to software projects. He is the creator of the technique *Earned Schedule*, which extracts schedule information from earned value data. Mr. Lipke is a graduate of the USA DoD course for Program Managers. He is a professional engineer with a master's degree in physics, and is a member of the physics honor society, Sigma Pi Sigma ($\Sigma\Pi\Sigma$). Lipke achieved distinguished academic honors with the selection to Phi Kappa Phi ($\Phi K\Phi$). During 2007 Mr. Lipke received the PMI Metrics Specific Interest Group Scholar Award. Also in 2007, he received the PMI Eric Jenett Award for Project Management Excellence for his leadership role and contribution to project management resulting from his creation of the Earned Schedule method. Mr. Lipke was selected for the 2010 Who's Who in the World. At the 2013 EVM Europe Conference, he received an award in recognition of the creation of Earned Schedule and its influence on project management, EVM, and schedule performance research. Most recently, the College of Performance Management announced that Mr. Lipke has been selected to receive the Driessnack Distinguished Service Award, their highest honor. Walt can be contacted at waltlipke@cox.net.

To see previous works by Walt Lipke, visit his [author showcase in the PM World Library](#).