A COMPARISON OF EARNED VALUE MANAGEMENT AND EARNED SCHEDULE AS SCHEDULE PREDICTORS ON DOD ACAT I PROGRAMS

By Capt Kevin T. Crumrine, USAF and Lt Col Jonathan D. Ritschel, PhD, USAF

ABSTRACT

Earned Schedule, since it was introduced by Walt Lipke in 2003, has been studied extensively in a variety of different fields and on programs of all sizes. However, Earned Schedule's viability as an extension of Earned Value Management (EVM) in Major Defense Acquisition Programs (MDAP) has yet to be effectively answered. The first aspect of this research explores the breadth of Earned Schedule's adoption by the System Program Offices (SPO) of the United States Air Force. The second phase of this research explores whether Earned Schedule is a more accurate and timely schedule predictor than the EVM technique currently employed by the United States Department of Defense (DoD). A series of five descriptive statistical tests were conducted on the Earned Value data for 64 Acquisition Category (ACAT) I MDAP's. This research finds Earned Schedule to be a more timely and accurate predictor than Earned Value Management.

BACKGROUND

One would be hard pressed to open the Wall Street Journal or the New York Times on any given day without encountering a number of articles related to the broad financial cuts the Department of Defense (DoD) will experience over the next decade. In the financially lean environment under which the DoD now operates, effective management of a program's cost and schedule performance has never been more vital. However, the DoD has long struggled with cost overruns and schedule delays on major acquisition projects. Earned Value Management (EVM) has been the premier method of program management and program cost forecasting within the DoD since its inception in the 1960s. EVM has long been hailed for its ability to identify to the decision-maker whether a program is going to be over cost or over schedule. However, EVMS's merit of forecasting schedule overages has been questioned in recent years (Lipke, 2003: 1). The predominant shortcoming of EVM is how it measures schedule performance: it quantifies schedule overages in terms of dollars ($), rather than in terms of time. This means of measurement is ambiguous and potentially confusing to program managers. To overcome this problem, a new schedule measurement technique, Earned Schedule (ES) was developed (Lipke, 2003: 1). Earned Schedule rectifies the ambiguities of traditional EVMS schedule analysis by expressing schedule measurements in terms of time. It has been argued that the critical development of ES provides program managers the predictive tool needed to determine project completion dates using EVM data.

Earned Schedule (ES) has been developed, and consequently studied significantly over the last decade, but a thorough and conclusive application to Acquisition Category I (ACAT I) Major Defense Acquisition Programs (MDAP) has yet to be accomplished. This paper researches whether ES is a more accurate predictor of schedule overruns than the current metrics used, whether it can predict schedule overruns earlier in the life of a program than the current schedule metrics, and how accurately ES can predict the final completion of a project on ACAT I DoD acquisition programs.

DATA SOURCE

The Office of the Under Secretary of Defense (OUSD) for Acquisition, Technology, and Logistics (AT&L) maintains the Defense Acquisition Management Information Retrieval (DAMIR) system. DAMIR is comprised of all contractor performance report (CPR) data for all Department of Defense (DoD) acquisition programs. The CPR "is a contractually required report, prepared by the contractor, containing performance information derived from the internal EVMS. [It] provides status of progress on the contract" (EVMIG 2006, 91). Within this CPR data are monthly and quarterly updates for all the Work Breakdown Structures (WBS) for a project.

This paper is an adaptation from Capt Crumrine's Masters thesis undertaken as part of a CPM sponsored research project. The full thesis (AFIT-ENV-13-M-36) is available from the AFIT website or CPM electronic library.
The article represents the views of the authors and not necessarily those of College of Performance Management. Sponsorship of a research project does not signify endorsement of the findings or conclusions by College of Performance Management.

This analysis focuses on ACAT 1 development and production contracts at the summary level (WBS 1). The programs comprising the dataset have completed their acquisition phase, and are either in their operational phase, or have been retired from the Air Force fleet. In addition to these programs, a "deep-dive" analysis will be completed on a single program: C-130J. While the development and production contracts mentioned above will be analyzed at the "program" level, the C-130J will be analyzed at the lower levels of the work breakdown structure. The purpose of the "deep dive" on the C-130J is to determine the efficacy of ES with respect to the critical path. Table 1 illustrates the complete dataset.

Platform Name	Number of Contracts	Number of Data Points
B-1	14	233
B-2	2	19
F-16	4	51
F-15	8	115
A-10	5	135
E-3	2	50
T-46	2	25
C-17	11	226
EF-111	2	36
AGM-131 (SRAM)	1	17
AGM-86 (ALCM)	8	74
AGM-65 (Maverick)	4	41
C-130J	1	65
Sum	64	1,087

Table 1: Platforms, Number of Contracts & Number of Data Points in Data Set

HYPOTHESIS TEST

The hypotheses for this research explore two distinctive avenues of program performance: timeliness and accuracy. The first facet attempts to answer whether using Earned Schedule for a program's schedule analysis would identify schedule overages with greater accuracy than the current Earned Value Management schedule analysis techniques. The second facet attempts to answer whether using Earned Schedule for a program's schedule analysis would identify schedule overages earlier in the life cycle of a program than the current Earned Value Management schedule analysis techniques.

Ho: Earned Schedule is not a more accurate and timely predictor of schedule overages than traditional Earned Value techniques.

Ha: Earned Schedule is a more accurate and timely predictor of schedule overages than traditional Earned Value techniques.

In failing to reject the null, the results of this analysis would indicate that using Earned Schedule rather than EVM for schedule management offers no additional benefit. If, however, the null is rejected, Earned Schedule proves to be a more accurate and timely predictor of schedule overages than traditional EVM methods.

METHODOLOGY AND RESULTS

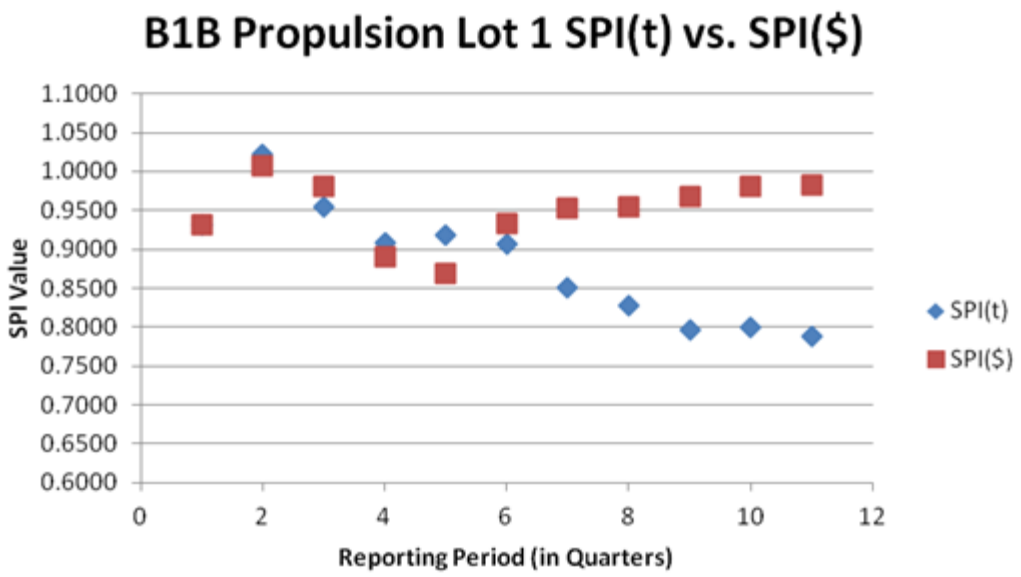
Prior to examining the accuracy and timeliness of Earned Schedule versus Earned Value Management, it must first be determined whether or not there is a statistical difference between these two methods. This is accomplished through a paired t-test. The paired t-test measures the mean difference between the ES and EVM numbers in the 1087 data points of Table 1. The null hypothesis of the paired t-test is that there is no difference between the methods. The resulting t-statistic of the paired t-test is -8.6231, with a p-value of 2.27467E-17 and a rejection of the null hypothesis. Therefore, it is determined that the difference between Earned Value Management and Earned Schedule is statistically significant. A further comparison of the timeliness and accuracy of the ES versus EVM methods is now warranted.

Now that it has been determined that there is a statistically significant difference between the results of the two schedule analysis methods, a series of five tests are conducted as a means of answering our two quantitative research questions on timeliness and accuracy. The

five descriptive statistical tests conducted on the earned value data yielded several useful measurements that support the value of using Earned Schedule.

The first test calculates the SPI($) for EVM and SPI(t) values for each data point and plots the points on a scatter plot to compare their behavior over the life of a program. Further, the average SPI values at six pre-determined completion points (20%/40%/50%/60%/80%/90%) are plotted to illustrate the behavior of the two methodologies over time.

The first test results illustrate that the SPI($) regresses to 1.0 over the final third of the program, while the SPI(t) offers a more realistic measure of the programs schedule. An illustration of a typical program is shown in Figure 1 for the B-1 Bomber.



Next, the average SPI values for both methods at the six predetermined program completion points were plotted on a scatter plot. The results of that test are shown in Figure 2 below.

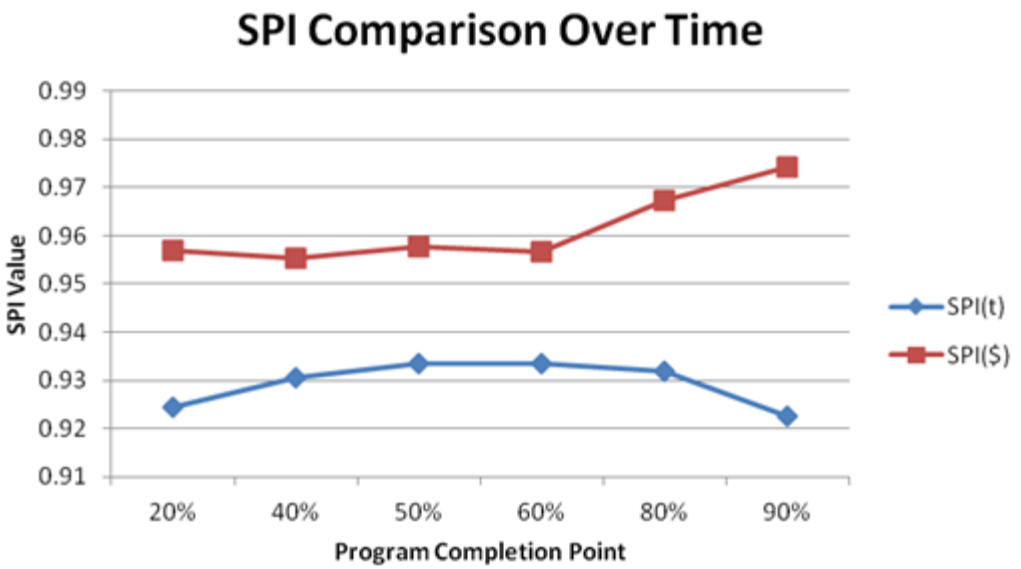


Figure 2: SPI Comparison Over Time

The SPI comparison in Figure 2 shows, at the portfolio level, what was expected from the SPI values. The SPI(t) and SPI($) values react with general consistency over the first two-thirds of the program, with the SPI values separated by only hundredths of a point. However, at the two-thirds completion point, the SPI values go in different directions: the SPI(t) value begins to decrease at a more rapid pace, while the SPI($) value regresses back towards a value of 1.0. We know that an SPI value of 1.0 indicates that a program is on schedule, but also that all

of the programs studied were delivered after their estimated completion dates. This supports previous research that states as a program progresses past the two-thirds completion point, SPI($) becomes an unreliable measure of program schedule while ES continues to provide value to the program manager (Lipke, 2003:1).

The second test analyzes the first time a program dropped below an SPI value of 0.90. The intent of this test is to determine whether Earned Value Management or Earned Schedule is an earlier detector of problems in meeting program schedule objectives. Because the literature lacked a conclusive threshold, the definition of a problem (SPI(t) or SPI($) < 0.90) is utilized.

Initial data analysis demonstrated that there are frequent occurrences where a program's SPI value drops below 0.90 early in a program and quickly recovers. This led to the potential for false conclusions. As a result, the total frequency of points below 0.90 for each method was also calculated.

On average, EVM identified a problem at the 18% completion point, while ES identified problems on average at the 30% completion point. However, because EVM failed to identify problems on 9 programs that ES did, we calculated the number of points below 0.90 for each method. EVM identified 12.4% of the 1,068 data points as below 0.90, while ES identified 20.3% of the data points. It was especially evident when studying the difference over time: at the 90% completion point, ES identified 20 programs as being "in trouble," while EVM identified only 1 program. A comparison of the number of programs with SPI values below 0.90 at the six pre-determined program completion points over time is shown in Figure 3 below.

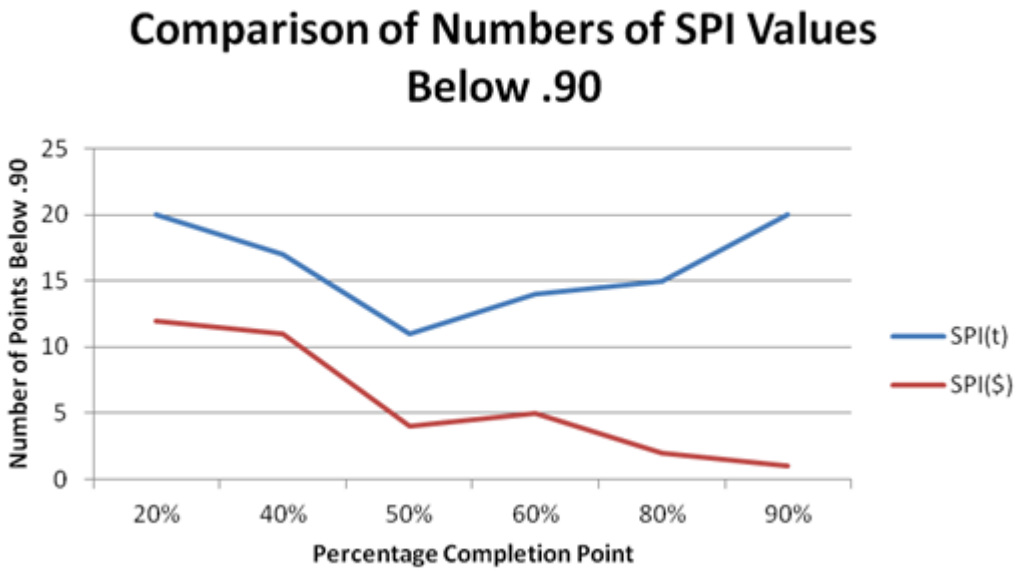


Figure 3: Comparison of Numbers of SPI Values Below .90

The third test examines the optimism and accuracy of the SPI values. Initial data analysis appeared to indicate that traditional schedule metrics are very optimistic: our test sought to test whether this was accurate. For each data point, whichever SPI value was higher was determined to be more optimistic:

EVM was more optimistic 59.4% of the time, while ES was more optimistic 35.1% of the time (5.5% of the data points had equal SPI values). Further, at the 90% completion point, EVM was more optimistic for 59 programs, while ES was more optimistic for 5 programs. A comparison of which method is more optimistic over time is illustrated graphically in Figure 4 below.

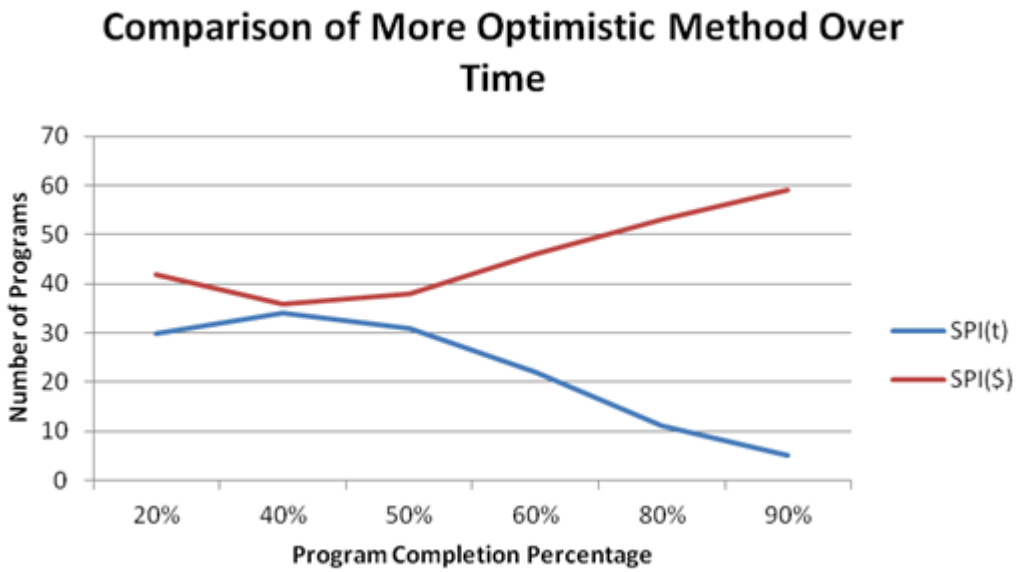


Figure 4: Comparison of More Optimistic Method Over Time

A similar methodology was used to determine which method was the more accurate: whichever method is closer to the final schedule result was determined to be the most accurate for that particular data point. EVM was more accurate 37.1% of the time, while ES was more accurate 57.4% of the time (again, 5.5% of the data points had equal SPI values). A comparison of which method is more accurate over time is illustrated graphically in Figure 5 below.

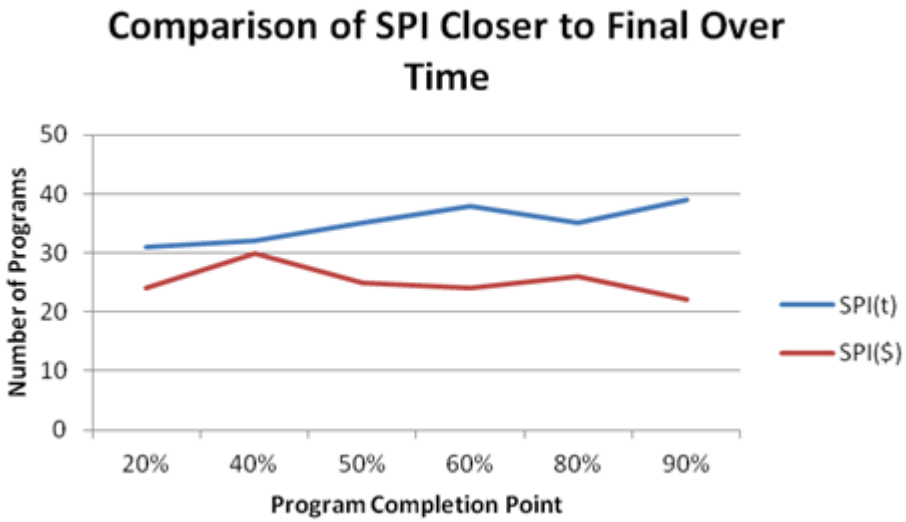


Figure 5: Comparison of SPI Closer to Final Over Time

The fourth test studies the To Complete Schedule Performance Index (TSPI), the schedule efficiency required for the remainder of the program to meet the original program schedule. The literature suggests that any TSPI value greater than 1.1 indicates that a program is unrecoverable and will not be delivered on time. We studied whether lowering this threshold to 1.05 or 1.01 would offer more value as a trigger to signal that a program is in trouble.

On average, a program reached a TSPI value at the 64% completion point. This doesn't notify the program manager until the program is 2/3 complete that there is a problem. Lowering the TSPI value to 1.05 identifies, on average that a problem exists at the 50.8% completion point. Finally, lowering the TSPI value to 1.01 identifies that, on average, a program is in trouble at the 24% completion point. Using a TSPI value of 1.10 identifies problems on all 64 programs studied. Using a TSPI value of 1.05 identifies problems on 63/64 programs. Finally, using a TSPI of 1.01 identifies problems on 61/64 programs. Figure 6 below demonstrates that using a TSPI value of 1.01 is far more useful at identifying early in the life of a program that issues exist. At the 40% completion point, 46 programs are identified as being "in trouble," using a TSPI value of 1.01,

while a TSPI value of 1.10 identifies only 10 programs. Lowering the TSPI threshold from 1.10 to 1.01 doesn't falsely identify programs as being in trouble when they truly aren't, and it also gives the program manager substantially more time to make resource allocation decisions to allow for on-time deliveries.

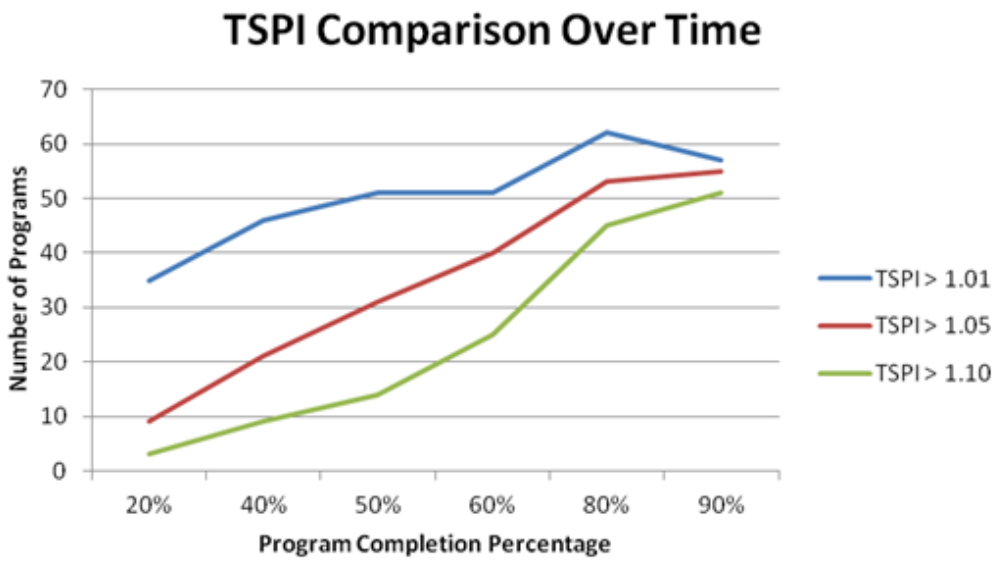


Figure 6: TSPI Comparison Over Time

The final test focused on the Schedule Variance (SV) calculations of the two methodologies: SV($) for EVM and SV(t) for ES. Our research did not focus on identifying the behavior of the two methods over the life of an average program: such observations have been well documented in the pertinent ES literature: rather, the intent of our research was to identify the SV inflection point, and more precisely, the percentage completion point at which SV($) begins its regression back to zero.

For each contract, the SV(t) and SV($) were calculated and graphed on a scatter plot. The scatter plot had two vertical axes: one measured in dollars to accommodate the EV calculation and one measured in time to accommodate the ES calculation. Figure 7 exhibits the scatter plot for a typical program, in this case the B1B Propulsion Lot 1 contract.

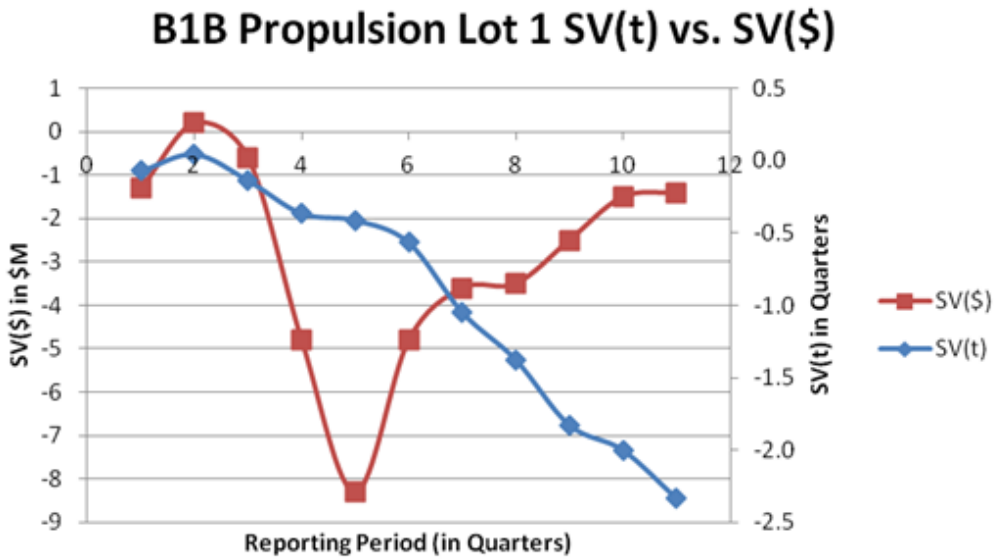


Figure 7: B1B Propulsion Lot 1 SV(t) vs. SV($)

For this analysis, the divergence point on each of the scatter plots is identified, and put in terms of program percentage completion. The divergence occurs when the SV($) is no longer growing and begins its regression towards zero. In Figure 7, the divergence is determined to occur during the 5th reporting period, which was the 45.45% completion point of the

program. The divergence points, in terms of percentage, were calculated for each of the programs. Of the 64 programs studied for this research, there were three where there was no discernible divergence point: ALCM Interface, F-15 Airframe Development and F-15 Engine Lot 7 Production. For our analysis, these three data points were removed. The mean divergence point was at 50.62% program completion. The results of this test indicate that at approximately the 50% program completion point, the SV($) metric provides little value to the decision maker. There are two advantages to using the Earned Schedule metric as a measure of schedule variance. The first, outlined extensively in the literature, is that Earned Schedule puts SV values in terms of time, rather than in terms of dollars. The second advantage, detailed in this research, is that Earned Schedule provides more accurate and realistic schedule variance indicators to the program manager over the second half of a program.

APPLICATION TO THE CRITICAL PATH

One criticism to Earned Schedule has been whether it is applicable when using the Critical Path Method (CPM). The literature, predominantly from Walt Lipke's 2006 paper titled *Applying Earned Schedule to Critical Path Analysis and More*, suggests that Earned Schedule is applicable to the Critical Path. This research investigated Earned Schedule's applicability to the Critical Path Method on the C130J Block 7.0 Upgrade program. The results show a fundamental disconnect between what Earned Value data is collected and the Critical Path data used by the DoD program office. Our example produced earned value data no deeper than the Work Breakdown Schedule (WBS) level 3 (ex: WBS Element 1.2.3). The Critical Path data is collected much deeper, as detailed as WBS level 7 (ex: WBS Element 1.2.3.4.5.6.7). This disconnect prevented us from conducting a detailed analysis. This does not necessarily suggest that Earned Schedule is inapplicable to the Critical Path Method in DoD, but conclusive research demands more detailed Earned Value data to be collected as part of the contract deliverables.

RESEARCH QUESTIONS ANSWERED

Our first question examined the extent that Earned Schedule is currently utilized in Air Force ACAT I acquisition programs. Through a series of surveys and interviews, we determined that no System Program Offices (SPO), regardless of DoD product center, currently use Earned Schedule as their primary schedule analysis tool. Of the seventeen program offices that responded to our inquiry, six program offices use Earned Schedule in some capacity. Those who use Earned Schedule use it exclusively as a secondary analysis tool, and as a cross-check to the results obtained from the Earned Value Management analysis. Based on the findings of our research, we believe that DoD ACAT I programs should embrace ES as a complementary tool to the CPM method that is predominately utilized.

Our second question examined whether Earned Schedule provides more accurate schedule predictions than EVM. We determined that Earned Schedule was a more accurate predictor of schedule than Earned Value Management. At the 50% completion point, the average SPI($) metric was .958, while the average SPI(t) metric was .934. Earned Schedule showed its value as a more accurate schedule predictor even later in the program: at the 90% completion point, the average SPI($) was .974, while the SPI(t) for the same data was .923. Another measure of how ES is more accurate than EVM was in how frequently each technique suggested the program was "in trouble." We determined that any SPI value below 0.90 meant the program was in trouble. Of the 1,087 points we studied, EVM indicated only 135 points (12.42%) were below 0.90, while ES disclosed that 220 points (20.24%) were determined to be in trouble (with no false positives). Another measure compared how ES predicted problems over time compared to EVM: at the 60% completion point, ES predicted 14 programs were in trouble, while EVM predicted 5 programs were in trouble. More telling, at the 90% completion point, ES indicated 20 programs were in trouble while EVM predicted only 1 would come in behind schedule. A further measure we used to determine which method was more accurate was the frequency of which method was more optimistic. The more optimistic measure was determined to be the higher of the two SPI values: the higher SPI value indicates the program is more likely to come in on time. Of our 1,087 data points, Earned Value Management was more optimistic for 646 (59.43%) of the data points, while ES was more optimistic for 381 (35.05%) of the data points. Earned Schedule proved to be more accurate later in the life of a program: EVM was more optimistic 38 times, compared to 31 programs, at the 50% completion point. However, at the 90% completion point, EVM was more optimistic on 59 programs, while Earned Schedule was more optimistic on only 5. Our series of tests confirmed that Earned Schedule is a more accurate schedule predictor than the Earned Value Management technique currently employed by the Department of Defense on Major Defense Acquisition Programs.

Our final question examined whether Earned Schedule provides timelier schedule predictions than traditional EVM methods. *We concluded that Earned Schedule was a timelier predictor of schedule overages than Earned Value Management*. While on average, Earned Value

Management first detected a problem in a program (SPI value below 0.90) at the 17.89% completion point as opposed to Earned Schedule detection at the 29.89% completion point, this failed to account for the 9 programs where EVM neglected to identify a problem existed while ES did. Further, at the 20% completion point, EVM only identified 12 programs that were "in trouble," while ES detected issues with 20 programs (with no false positives). A comparison of the SPI Closer to the Final Schedule concluded that Earned Schedule was closer to the final result more frequently than Earned Value Management as early as the 20% program completion point. The further the program progressed, the more frequently the SPI(t) was closer to the final schedule. Finally, it was determined that the SV(t) and SV($) had an average divergence point of 50.62%. This conclusion dictates that the SV($) metric is comparatively useless over the final half of the program, while Earned Schedule provides meaningful information over the entire life of a program.

FUTURE RESEARCH

Our research was conducted on system platforms from a single product center, and was accomplished at a general level. We recommend conducting parallel research on the other two major system platforms: Electronic systems, as well as Space and Missiles systems. There are several differences between the particular platforms that could lead to dissimilar results than what we found: the size of the programs, the length of the acquisition process, the integration of Commercial Off The Shelf (COTS) products, and the level of security classification. Another future opportunity builds off of this research: A comparison of the different classes of aeronautical platforms, researching if there is a significant difference between implementing Earned Schedule versus Earned Value Management on Fighter Aircraft, Bomber Aircraft, Cargo Aircraft, Unmanned Aircraft, and Rotary Wing Aircraft. Again, there are differences between these platforms based on technological sophistication and number of units produced. An additional proposal for future research would be very similar to the comparison of different classes of aeronautical platforms: a comparison of how well Earned Schedule performs as a schedule predictor versus Earned Value Management for the different prime contractors used by the DoD.

REFERENCES

Defense Contract Management Agency. . Department of Defense Earned Value Management Implementation Guide. October 2006. Alexandria, VA: DCMA, 2006. 10 Sept 2012 http://guidebook.dcma.mil/79/EVMIG.doc

Lipke, Walt. "Applying Earned Schedule to the Critical Path and More," The Measurable News, Fall 2006

Lipke, Walt. "Schedule is Different," The Measurable News, March & Summer 2003

ABOUT THE AUTHORS

Kevin T. Crumrine, Captain, USAF. Captain Crumrine is a cost analyst at the Air Force Life Cycle Management Center, Hanscom AFB, MA. Prior to his assignment to Hanscom AFB, he graduated from the Air Force Institute of Technology in March 2013 with a M.S. in Cost Analysis. Kevin received his commission upon graduation from the United States Air Force Academy in 2008, with a B.S. in Management. From there, he was assigned as the Financial Analysis Flight Commander for the 2nd Bomb Wing at Barksdale AFB, LA. Email: kevin.crumrine@gmail.com

Jonathan D. Ritschel, Lieutenant Colonel, USAF. LtCol Ritschel is the Director, Cost Analysis Graduate Program and Assistant Professor in the Systems Engineering and Management Department at the Air Force Institute of Technology. He received his BBA. in Accountancy from the University of Notre Dame, an MS in Cost Analysis from the Air Force Institute of Technology, and a PhD. in Economics from George Mason University. His work has been published in various journals such as Economic Papers, Journal of Public Procurement, Journal of Cost Analysis and Parametrics, and Air Force Journal of Logistics. Email: jonathan.ritschel@afit.edu