JOURNAL OF PUBLIC PROCUREMENT, VOLUME 15, ISSUE 3, 341-364 FALL 2015

STABILITY PROPERTIES IN DEPARTMENT OF DEFENSE CONTRACTS: ANSWERING THE CONTROVERSY

Jacob L. Petter, Jonathan D. Ritschel and Edward D. White*

ABSTRACT. Delineating where stability occurs in a contract provides the window of opportunity for procurement officials to positively affect cost and schedule outcomes. While the concept of a Cost Performance Index (CPI) "stability rule" has been routinely cited by Earned Value Management (EVM) authors since the early 1990's, more recent research questions the veracity of this stability rule. This paper resolves the controversy by demonstrating that the definition of stability matters. We find a morphing of the stability definition over time, with three separate definitions permeating the literature. Next, an analysis of Department of Defense contracts for both cost and schedule stability properties finds that the veracity of the stability rule is intricately tied to the definition used.

INTRODUCTION

The United States Department of Defense (DoD) oversees a portfolio of 86 Major Defense Acquisition Programs (MDAPs) valued at over $1.6 trillion (GAO, 2013). Successfully managing this is a monumental task even under the best economic conditions. In a

---

** Jacob L. Petter, MS, is a cost analyst at the Air Force Cost Analysis Agency. His research interests include public procurement, public budgeting, and acquisition cost estimating. Jonathan D. Ritschel, Ph.D., is an Assistant Professor, Department of Systems Engineering and Management, the Air Force Institute of Technology. His research interests include public choice, public procurement, and economic institutional analysis. Edward D. White, Ph.D., is a professor, Department of Mathematics and Statistics, the Air Force Institute of Technology. His research interests include design of experiments, growth curves, linear and nonlinear regression, and categorical data analysis.* *The views expressed in this article are those of the authors and do not necessarily reflect the view of the Department of Defense or any of its agencies.*

constrained fiscal environment where future DoD budgets are projected to decline, key decision makers must rely more heavily on measurements that can accurately predict if a program will be successful in adhering to the budget or become a financial catastrophe. Thus, if decision makers can conclude early on whether or not the program will succeed, better financial decision are made.

Within the Earned Value Management (EVM) system, there is a rule of thumb about a cost performance measurement that does just that. The Cost Performance Index (CPI) "stability rule" states that the cumulative CPI does not change more than a certain amount once the contract is 20% complete. Research on DoD MDAPs dating back to 1992 supports this claim (Christensen & Payne, 1992; Christensen & Heise, 1993; Christensen & Templin, 2002). The literature review reveals, however, that each of these studies defines stability somewhat differently. Specifically, stability in previous studies is defined through either a *range* or *interval* interpretation. There is a key element that differentiates between these two approaches. The *range* interpretations of stability calculate over the entire span of a contract's life through analysis of minimum and maximum CPIs over the range. In contrast, the *interval* interpretations of stability examine CPI only at discrete points in time compared to the CPI final. A more detailed explanation of the stability definition and differences is provided in a subsequent section.

Additionally, the stability rule itself has been questioned. Recent research claims contradictory evidence that denies CPI stability as generalized within the DoD portfolio (Henderson & Zwikael, 2008). Because of the conflicting results and differing stability definitions, the question remains whether CPI stability exists today in the DoD. The answer has implications for resource allocation in the public arena.

Although the focus of this paper is on United States DoD contracts, there are myriad other areas of relevance. EVM is utilized by procurement official in a multitude of countries such as Australia, Brazil, Japan and Sweden (Antvik, 2001; Marshall, 2008). Europe has become a leader in EVM application, with annual EVM-Europe conferences and significant research originating from Ghent University in Belgium (Vandevoorde & Vanhoucke, 2006). Additionally, EVM is widely utilized in both the public and private sector. EVM usage is found in a wide array of project types from construction to software development (Bhosekar & Vyas, 2012; Henderson & Zwikael, 2008). Therefore, while this analysis is limited to the United States, and specifically the Department of Defense, the insights are relevant to a wide range of other countries and industries.

While analyses of cost stability properties (i.e. stability of CPI) have remained in the forefront, a parallel analysis of schedule stability remains unexplored. What about possible stability in schedule performance measurements, such as the Schedule Performance Index (SPI) or Earned Schedule's Schedule Performance Index (SPI(t))? Thus, this research re-examines CPI stability with an updated portfolio of DoD data, and then conducts the first-ever schedule stability analysis.

EARNED-VALUE MANAGEMENT IN THE DOD

Earned-value management (EVM) in the DoD acquisition community is integral to the acquisition program management value chain. It provides program managers accurate and timely insight into the cost, schedule, and performance of DoD weapon system programs. EVM originated from the directive-imposed Cost/Schedule Control Systems Criteria (C/SCSC) that the DoD set as the standard for all programs in 1967 (Fleming & Koppelman, 1998). The C/SCSC consisted of 35 criteria that contractors were required to meet when under a contract with the US government. These criteria were subsequently updated by the National Security Industrial Association (NSIA) in 1995 to 32 criteria and became EVM. The DoD endorsed and implemented this criteria in 1996 (Fleming & Koppelman, 1998). A summary of key EVM terms and concepts is presented in Table 1.

Two primary EVM measurements that together show the overall performance (cost and schedule) of a program are the efficiency indices: Cost Performance Index (CPI) and Schedule Performance Index (SPI). The CPI is a ratio between the Budgeted Cost for Work Performed (BCWP) and the Actual Cost of Work Performed (ACWP) (DAU, 2013). It indicates the value of every dollar of work accomplished. For example, a CPI of 0.98 means the program is receiving 98 cents of budgeted value for every dollar spent. The CPI is calculated with either current or cumulative data. For the purposes of this research, CPI refers to the cumulative value.

TABLE 1
Earned-Value Management Concepts

EVM Measurement	Acronym	Definition	Formula
Budgeted Cost for Work Performed	BCWP	The earned value, how much budgeted cost the project has gained within a given time period	Sum of the budgeted cost of all completed work packages
Actual Cost of Work Performed	ACWP	The actual cost of the completed work packages within a given time period	Sum of actual costs of all completed work packages
Budgeted Cost for Work Scheduled	BCWS	The planned value, how much budgeted value the project is scheduled to accomplish within a given time period	Sum of the budged cost of all work packages scheduled
Cost Perfor-mance Index	CPI	Cost efficiency of a project	BCWP/ACWP
Schedule Performance Index	SPI	Schedule efficiency of a project	BCWP/BCWS
Schedule Performance Index – Time	SPI(t)	Schedule efficiency of a project	Earned Schedule/Actual Time
Budget at Complete	BAC	Planned total cost of a project	Sum of all BCWS
Estimate at Complete	EAC	Forecasted total cost of a project	Various formulas
To Complete Performance Index	TCPI	Projects what the CPI will be for the remainder of the project to meet the BAC	Various formulas

SPI is the second EVM efficiency index. It is a ratio between the BCWP and the Budgeted Cost for Work Scheduled (BCWS) (DAU, 2013). EVM's SPI, however, has two well-known drawbacks for program managers. First, the SPI ratio is comprised of two budgeted values that converge to one another at contract completion. As a result, the SPI equals 1 at the end of the contract regardless of actual schedule performance. Research has demonstrated that SPI is no longer useful in measuring schedule performance once the contract is approximately 67% complete (Lipke, 2003). Second, SPI is a ratio of two dollar values, rather than units of time.

In response to these shortcomings, a method called Earned Schedule (ES) was developed (Lipke, 2003). The primary purpose of ES is to provide time-based measures of schedule utilizing EVM data. A branch off of EVM, ES is useful the entire span of a contract, and has its own schedule performance index utilizing time rather than budget costs. This measurement is the SPI(time), or simply SPI(t). SPI(t) uses the time increments associated with the BCWP and BCWS to calculate performance. See Lipke (2003) for a technical description of ES. In the DoD, ES has not become a primary technique, but is gaining popularity in Air Force acquisition program offices as a crosscheck methodology (Crumrine and Ritschel, 2013). Because of its accuracy over the entire lifetime of a contract, SPI(t) is the metric used in this research.

WHY STABILITY?

There are multiple benefits to DoD Program Managers (PMs) if efficiency indices (CPI and SPI(t)) are found to possess stability properties. First, stability provides a better forecasting or estimating ability and therefore affects investment decisions. For example, CPI is part of the calculation for the projected Estimate at Complete (EAC) of a project/program. Therefore a stable CPI would provide better EACs due to reduced variation. The enhanced forecasting allows PMs to avoid the mistake of committing more funds to a failing project, which can be very costly to all stakeholders involved with that program (Christensen, 1996). Second, the CPI serves as a “benchmark” to the To Complete Performance Index (TCPI), which calculates the CPI needed for the remainder of the contract’s life in order to meet the Budget At Complete (BAC) (Christensen and Heise, 1993). If the TCPI is significantly higher than the CPI, the TCPI may be unattainable when the CPI is stable. Therefore, the program will not be able to reach the BAC and will most likely finish over budget. Determining the veracity of CPI stability gives confidence to this conclusion and enhances financial management decisions. Third, a stable performance index is evidence that the contractor’s management system is working. Stable performance indicates that there are no large variances unaccounted for or unaddressed by the management of that contract (Christensen and Heise, 1993). While the CPI is the subject of all these referenced benefits, stability in either of the efficiency indices will inform better decisions for public procurement professionals.

LITERATURE REVIEW: DEFINING STABILITY

Many training guides, handbooks, and authors in the EVM community quote the CPI "stability rule." Thus, it is imperative to clearly define the term "stability." While it may appear trivial on the surface, a literature review of the CPI definition of stability finds that the definition has morphed over time.

The seminal stability article was published by Christensen and Payne in 1992. They examined 26 Contract Performance Reports (CPRs) from seven DoD aircraft procurement contracts. Utilizing two methods, they investigated stability properties at the contract's 50 percent completion point. The first method was a 0.2 *range* of minimum and maximum CPIs from the fifty percent completion point to the end of the contract. The second method was an *interval* of "plus and minus 10 percent of the CPI computed at the fifty percent point" (Christensen & Payne, 1992, p. 2). Although the focus was to determine stability from 50 percent complete onward, their results with the *range* method indicated stability occurred at the 20 percent completion point. The *interval* method resulted in all the contracts being stable at the 50 percent completion point but not earlier.

As a result of Christensen and Payne's preliminary findings, Christensen and Heise (1993) utilized the *range* definition in an expanded DoD dataset. They found that "based on an analysis of 155 contracts from the DAES database, the cumulative CPI was stable from the 20 percent completion point with a 95 percent confidence interval" (Christensen & Heise, 1993, p. 5). This result is the origin of what became commonly referred to as the "*CPI stability rule*." It is important to note that Christensen and Heise did not claim their finding to be a "stability rule" nor generalizable beyond the data set. Rather, subsequent authors and EVM practitioners cited their work with this moniker.

The definition of stability again morphed in 2002 with the work of Christensen and Templin. Examining 240 DoD contracts, they tested a general rule of thumb, "the cumulative [CPI] will not change by more than 0.10 from its value at the 20 percent completion point, and in most cases it only worsens" (Christen and Templin, 2002: 1). They found, with few exceptions, that the difference between the final CPI and the CPI at 20% complete was less than or equal to 0.10 (Christensen & Templin, 2002).

The veracity of the CPI stability rule was questioned by Kym Henderson and Ofer Zwikael (2008) in a study of small non-DoD projects. Their dataset consisted of forty-five projects dealing with information technology and construction in the United Kingdom, Israel, and Australia. They used Christensen and Templin's *interval* definition of stability as the difference between the final cumulative CPI and the cumulative CPI at the 20% complete point being no more than plus or minus 0.1. Their analysis concluded that the contracts did not stabilize by the 20% complete point, but in fact, "the stability is usually achieved very late in the project life cycle, often later than 80 percent complete for projects in these samples" (Henderson & Zwikael, 2008, p. 9). Henderson and Zwikael then turned their attention to DoD CPI stability findings. Through visual examining of CPI scatterplots from secondary DoD data (an unpublished study by Michael Popp, 1996), they concluded that the "widely reported CPI stability rule cannot be generalized to all projects utilizing the EVM method or even within the DoD project portfolio," (Henderson & Zwikael, 2008, p. 7; see Abba [2008] for a critical critique of this work).

Clearly, the definition and interpretation of "stability" has changed over time. Table 2 lists some of the literature citing or examining the stability rule beginning with the seminal work. Note that even the more recent citations have varying definitions/interpretations of stability.

TABLE 2
Stability Rule Citations

Author	Stability Interpretation
Christensen & Payne, 1992	"range of less than 0.20" and "plus or minus 10 percent of the CPI"
Christensen & Heise, 1993	"range being less than 0.2"
Christensen, 1996	"once a program is twenty percent completed, the cumulative CPI does not change by more than ten percent"

TABLE 2 (Continued)

Author	Stability Interpretation
Christensen, 1999	"cum CPI does not change by more than 10 percent from its value at 20 percent complete"
Christensen & Templin, 2002 (redefined stability)	"cumulative CPI will not change by more than 0.1 from its value at the 20 percent completion point"
Lipke, 2005	"the final value of the CPI does not vary by more than 0.1 from the CPI when the project is 20 percent complete"
Henderson & Zwikael, 2008	"within 0.10 of its value when the project is 20 percent complete"
Fleming & Koppelman, 2008	"plus or minus 10 percent"
GAO, 2009	"Once a program is 20 percent complete, the cumulative CPI does not vary much from its value (less than 10 percent)"
SCEA, 2010	"Cum CPI will not change more than 10% from the value at the 20% complete point in time"

Although Table 2 clearly shows that even today there are multiple interpretations of what stability means, these stability definitions can be summarized under three broad categories: *range* definition, *absolute interval* definition, and *relative interval* definition. These three categorical definitions are the basis for the analysis of this paper (See Table 3).

TABLE 3
Stability Definitions

Definition Name	Stability Definition	Stability Literature Sources
Range	When the difference between the maximum and minimum SPI(t) (or CPI) between a specific percent complete and the final point is less than 0.2.	Christensen & Payne (1992); Christensen & Heise (1993)
Absolute Interval	When the final SPI(t) (or CPI) is within 0.10 of the SPI(t) (or CPI) at a specific percent complete.	Christensen & Templin (2002); Lipke (2005); Henderson & Zwikael (2008)

TABLE 3 (Continued)

Definition Name	Stability Definition	Stability Literature Sources
Relative Interval	When the difference between the final SPI(t) (or CPI) and the SPI(t) (or CPI) of a specific percent complete is less than or equal to plus or minus 10% of the SPI(t) (or CPI) at the specific percent complete.	Christensen (1996); Flemming & Koppelman (2008); GAO (2009); SCEA (2010)

While there is much research and literature about CPI stability, there is little research on possible SPI(t) stability. Henderson and Zwikael's (2008) research effort is the lone study to date that examined SPI(t) data for stability; however they did not analyze DoD data. They found that SPI(t) stability was not achieved in their sample of 45 international projects until very late in the project.

DATA

The dataset for this research is from the Defense Acquisition Executive Summary (DAES) and Defense Cost and Resources Center (DCARC) databases. Only those contracts associated with Acquisition Category I (ACAT I) programs are included. The period of analysis spans 25 years from 1987 to 2012. Any ACAT 1 contract that is no more than 10% complete by the start of 1987 and at least 85% complete by the end of 2012 (percent complete is defined as cumulative BCWP divided by the final BAC) is included. These thresholds ensure complete contract data, which is essential for examining stability properties.

Consistent with previous research (Christensen and Heise, 1993; Christensen and Templin, 2002), all contracts with Over-Target-Baselines (OTBs) are removed from the dataset. With an OTB (and subsequent re-baseline), the budget of a contract changes. This change causes all the EVM budgeted measurements to restart, including the cumulative measurements. Table 4 displays a summary of the dataset before and after the OTBs were removed and the time window/percent complete requirements were enforced. The final dataset includes 209 development and production contracts from the Air Force, Navy, and Army.

TABLE 4
Dataset Characteristics

Category	Total Contracts
Preliminary Total (before percent complete and OTB cuts)	822
Contracts with OTBs	165
Does not meet the time window/percent complete requirements	447
Final Dataset for analysis	209

METHODOLOGY

The methodology of this research includes two parts: a variance analysis investigating CPI and SPI(t) stability in all contracts and a comparison analysis examining CPI and SPI(t) stability differences by Service, Contract Type, Life-Cycle Phase, Contract Length, and Platform Type.

Variance Analysis

Variance analysis is conducted on CPI and SPI(t) for every contract to determine when it stabilizes. The analysis is executed on a contract's CPI and SPI(t) from the 10 percent complete point to 85 percent complete point in increments of five. Three separate analyses are completed for both CPI and SPI(t): one analysis for each definition in Table 3.

First Analysis: Range Definition of Stability

First, stability is defined as when the difference between the maximum and minimum CPI (or SPI(t)) between a specific percent completion and the final point is less than 0.2. This is the *range* definition of stability. The test for this definition is:

Stable if: $|CPImax_{x\%} - CPImin_{x\%}| \leq 0.20$
Unstable if: $|CPImax_{x\%} - CPImin_{x\%}| > 0.20$

$CPImax_{x\%}$ equals the maximum CPI from the *X*% complete point to the final CPI of the contract, and $CPImin_{x\%}$ equals the minimum CPI from the *X*% complete point to the final CPI of the contract. Percent complete (*X*%) is defined from 10 to 85 percent complete in increments of five. The calculated range at each percentage complete increment is recorded and the contract is categorized as stable or unstable. This formula tests for CPI stability. SPI(t) stability is calculated in the same manner, by replacing CPI with SPI(t).

Second Analysis: Absolute Interval Definition of Stability

For the second analysis, stability is defined as when the final CPI (or SPI(t)) is within 0.10 of the CPI (or SPI(t)) at a specific percent complete. This is the *absolute interval* definition of stability. The test for this analysis is:

Stable if: $|\text{CPI}(\textit{final}) - \text{CPI}(X\%)| \leq 0.10$
Unstable if: $|\text{CPI}(\textit{final}) - \text{CPI}(X\%)| > 0.10$

CPI(*final*) is the final CPI of the contract, and CPI(*X*%) is the CPI when the contract is *X*% complete. Percent complete (*X*%) is defined from 10 to 85 in increments of five. For example, the contract is stable from 20% complete if the difference between the final CPI and the CPI at 20% complete is less than plus or minus 0.10 (in this example, *X* equals 20%). Additionally, all subsequent percent complete increments after 20% complete must be stable as well for the contract to be considered stable from 20% complete. If the difference is greater than 0.10 or the contract becomes unstable after 20% complete, it is not considered stable at 20% complete. SPI(t) stability is calculated in the same manner, by replacing CPI with SPI(t).

Third Analysis: Relative Interval Definition of Stability

For the third analysis, stability is defined as when the difference between the final CPI and the CPI of a specific percent complete is less than or equal to plus or minus 10% of the CPI at the specific percent complete. For instance, a contract with a CPI at 20% complete of 0.8 stabilizes from that point if the final CPI is less than or equal to plus or minus 0.08 (ten percent of 0.8) away from 0.8. This is the *relative interval* definition of stability. The test is:

Stable if: $|\text{CPI}(\textit{final}) - \text{CPI}(X\%)| \leq 0.10 * \text{CPI}(X\%)$
Unstable if: $|\text{CPI}(\textit{final}) - \text{CPI}(X\%)| > 0.10 * \text{CPI}(X\%)$

CPI(*final*) is the final CPI of the contract, and CPI(*X*%) is the CPI when the contract is *X*% complete. Percent complete (*X*%) is defined from 10 to 85 in increments of five. Similar to the *absolute interval* definition, the contract must become stable at *X*% and remain stable at all subsequent percent complete increments to be considered stable at *X*%. SPI(t) stability is calculated in the same manner, by replacing CPI with SPI(t).

Comparison Analysis

After determining overall CPI and SPI(t) stability points, categorical comparisons are performed for both the *range* and *interval* definitions. Separate comparisons are not needed for the *absolute* interval and *relative* interval definitions as the interval calculation itself is the same – it is the criteria compared to the interval that are unique. The range is calculated as:

$$\text{Range} = |CPImax_{x\%} - CPImin_{x\%}|$$

$CPImax_{x\%}$ equals the maximum CPI from the *X*% complete point to the final CPI of the contract, and $CPImin_{x\%}$ equals the minimum CPI from the *X*% complete point to the final CPI of the contract. The same formula is utilized for SPI(t) by replacing CPI with SPI(t). For both interval definitions of stability, the interval is calculated as:

$$\text{Interval} = |CPI(final) - CPI(X\%)|$$

CPI(*final*) is the final CPI of the contract, and CPI(*X*%) is the CPI when the contract is X% complete. Percent complete (X%) is defined from 10 to 85 in increments of five. Again, SPI(t) is calculated by replacing CPI with SPI(t).

Four different comparisons are executed: Service, Contract Type, Life-Cycle Phase, and Platform. Table 5 lists the different services, contract types, life-cycle phases, and platforms analyzed.

TABLE 5
Categories for Comparison Analysis

Categories			
Services	**Contract Types**	**Life-cycle Phases**	**Platforms**
Air Force	Fixed Price	Development	Aircraft System
Army	Cost Plus	Production	Electronic/Automated System

TABLE 5 (Continued)

Categories			
Services	**Contract Types**	**Life-cycle Phases**	**Platforms**
Navy			Missile System
			Ordnance System
			Ship System
			Space System
			Surface Vehicle System

For each of these categorical comparisons, median values are compared with overall Kruskal-Wallis (KW) tests (when comparing more than two groups) and individual Mann-Whitney (MW) tests (when comparing two groups). For the individual tests, the hypotheses are (Hayter, 2007):

KW - Ho: $\Delta_X = \Delta_Y = \cdots = \Delta_Z$
Ha: At least one median is not equal to the rest.

MW - Ho: $\Delta_X = \Delta_Y$
Ha: $\Delta_X \neq \Delta_Y$

Where:
Δ equals the median of the ranges (from the first stability analysis) or interval (from the second stability analysis) for the specific percent complete (values of 10% to 85%, in increments of five).
X and *Y* are different groups in each comparison (for KW, *Z* represents however many groups are in the comparison). For example, when comparing Services, *X* and *Y* (for the MW test) are defined as Air Force and Navy, Air Force and Army, and Navy and Army respectively (three different tests).

First, the Kruskal-Wallis test is conducted at alpha of 0.05 to test whether there is a difference between the three groups. If significant results occur, three Mann-Whitney tests are conducted with alphas of 0.05/3 or 0.0167 to compare each pair of services, in accordance with the Bonferroni method of multiple simultaneous comparisons (Kutner, Nachtsheim & Neter, 2004). If the MW's null hypothesis is not rejected, then it is concluded that there is no difference in the median range or interval of the groups *X* and *Y*. If the MW's null hypothesis is rejected, it indicates there is a statistically significant difference in the medians. The group with a smaller median range or interval is considered to be more stable using that definition.

RESULTS: VARIANCE ANALYSIS

Table 6 displays the results of the CPI variance analysis. The percentage of contracts that possess stable CPIs at specific percent complete points is provided for all *three definitions* of stability utilized throughout this research. Bolded values in the table indicate when the stable percentage reaches 70, 80, and 90 percent. These stability percentages (70, 80, and 90) are highlighted because the literature review indicates they may be important milestones in determining stability tendencies. For example, from these numbers in Table 6, one can state at least 90 percent of the contracts have stable CPIs from 35 percent complete, using the range stability definition.

TABLE 6
CPI Stability Percentages

CPI Summary: Percentage of stable contracts																
Percent Complete	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**
Range Stability	**73**	79	**83**	84	88	**91**	93	94	95	96	98	98	99	99	99	99
Absolute Interval Stability	40	51	56	60	62	67	69	**72**	76	**81**	84	86	89	**92**	96	99
Relative Interval Stability	40	51	55	58	60	64	68	69	**73**	78	**82**	84	87	**90**	94	98

Depending on the definition of stability, the CPI results in Table 6 both support and contradict the "stability rule" within the DoD and the findings of earlier research. With the range definition of stability, 83 percent of the contracts possess a stable CPI when the contract is 20 percent complete, which is similar to the 86 percent of contracts being stable at the 20% completion point from Christensen and Heise's research. However, with either of the *interval* stability definitions, only about 55 percent of the contracts possess stable CPIs at 20 percent complete. This contradicts the interpretation of the stability rule that states the "cumulative CPI will not change by more than 0.1 from its value at the 20 percent completion point" (Christensen & Templin, 2002: 5). *The primary insight is that CPI* *stability behaves differently depending on the stability definition used.*

Table 7 displays the results of the SPI(t) variance analysis for all three definitions of stability. Utilizing the *range* definition, SPI(t) stability is found to be very similar to CPI stability. The 80 percent contract stability threshold is reached for both CPI and SPI(t) stability at the 20 percent completion point. In contrast, using either *interval* definition, the SPI(t) is not stable until much later than the corresponding CPI. The SPI(t) interval definition obtains stability for 80 percent of the contracts once the contracts are 75 percent complete, while the CPI interval definitions were much earlier at 55 and 60 percent complete.

TABLE 7
SPI(t) Stability Percentages

SPI(t) Summary- Percentage of stable contracts																
Percent Complete	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**
Range Stability	**72**	78	**82**	83	83	86	88	88	89	**90**	91	92	94	95	96	97
Absolute Interval Stability	41	52	58	59	62	66	68	69	**70**	73	75	78	79	**84**	87	89
Relative Interval Stability	38	50	54	55	57	61	62	64	65	67	69	**74**	76	**80**	85	88

RESULTS: COMPARATIVE ANALYSIS

Services

First, an overall Kruskal-Wallis test determines if there is a difference between the three services: Air Force, Army, and Navy. If the overall Kruskal-Wallis has significant results, a Mann-Whitney test is utilized to compare pairs of services' median ranges and intervals (Air Force v Navy, Air Force v Army, Navy v Army). If a Service has a smaller median CPI or SPI(t) range or interval than the Service it is compared to, it's CPI or SPI(t) is more stable when using that particular stability definition.

Table 8 summarizes the results. First, note that SPI(t) is not included in the table. There were only three instances in all the tests where SPI(t) was different. Thus, the conclusion is that there are no schedule stability differences between the Services.

However, there are statistically significant CPI stability differences between Services. In Table 8, an R indicates differences with the range definition; I indicates differences with the interval definition; B indicates differences with both the range and interval definition. Cells with "-" signify no significant results.

TABLE 8
Comparing CPIs by Service

Mann-Whitney tests results for Service																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
AF vs Army	-	-	-	-	-	R	R	-	-	-	B	-	-	-	I	-
Navy vs Army	R	R	R	R	R	R	B	R	R	R	B	-	-	-	-	-
AF vs Navy	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

For each increment from 10 percent complete to 60 percent complete, Navy has a smaller median CPI range than Army. The Air Force v Army tests only have a few significant differences, with the Air Force having smaller medians. Therefore, overall, Army either has statistically the same median or a greater median than the other two services. This is evidence that Air Force and Navy contracts' CPIs may be more stable than Army's, but not necessarily in every instance. It can definitely be stated, however, that Army contracts' CPIs are not more stable using either stability definition. This is consistent with earlier research findings by Christensen and Templin in 2002. Using the absolute interval definition, they found that from 20% complete Army's mean interval was larger than AF and Navy's mean intervals (Christensen & Templin, 2002).

Contract Type

Next, comparisons are made by contract type. The two categories compared are Cost Plus (CP) and Fixed Price (FP). Mann-Whitney tests at an alpha of 0.05 compare the medians at every percent complete (in increments of five) from 10% to 85%.

There are no significant results from the Mann-Whitney tests on CPIs of CP and FP contracts by either stability definition. Therefore, there is no difference in CPI stability between the two contract types. This finding is consistent with past research. Using the absolute interval definition of stability, Christensen and Templin (2002) found the mean intervals of CP contracts to be very similar to FP contracts. Christensen and Heise (1993) found only a slight difference (of just 1 percent) in the mean stability points of CP and FP contracts, using the range definition.

In contrast, differences are found in SPI(t) stabilities between contract types. Table 9 displays the results, where R, I, and B indicate range, interval or both.

TABLE 9
Comparing SPI(t) Stability by Contract Type

Mann-Whitney test results for SPI(t) - CP v FP																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
CP v FP	B	B	R	R	B	B	R	R	R	B	B	B	B	B	B	R

For all statistically significant tests, FP contracts have a greater median than CP. Thus, the SPI(t) of CP contracts tend to be more stable than FP contracts. These SPI(t) stability results are surprising because CP contracts are typically utilized when there is more uncertainty involved in the contract. A typical example of a CP contract would be the development effort for a new bomber. This type of contract would logically be thought to have more variation in schedule performance, but the results found here are contrary. One possible explanation is that the contractors may add contract change proposals or an engineering change to an FP program when the contractor is losing money. By attempting to increase the scope and receive more money, the schedule suffers, as there will be no money for overtime or to hire more personnel in an attempt to catch up. Also, contractors may use "other techniques [including] negotiating meaninglessly general statements of work, or agreeing to successive, after-the-fact, incremental fixed-price contracts that simply reimburse contractors for work already performed" which will ultimately worsen the performance of FP contracts (Singer, 1982, p. 11).

Life-Cycle Phase

The third comparison divides the contracts into their life-cycle phase: Production or Development. There were no operating & support (O&S) contracts considered in this analysis. Mann-Whitney tests at an alpha of 0.05 determine if differences exist in CPI or SPI(t) median values.

There are no SPI(t) statistically significant results. CPI, however, is found to be statistically significant from 10 percent complete to 60 percent complete. Over this range, the development contracts have a higher median than the production contracts. See Table 10 where R indicates differences with the range definition.

TABLE 10
Comparing CPI Ranges by Life-cycle Phase

Mann-Whitney test results for CPI																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Ranges	R	R	R	R	R	R	R	R	R	R	R	-	-	-	-	-

Production contracts are found to have more stable CPIs than development contracts until very late in the program. Intuitively, production contracts would be expected to be more stable as they are typically an iterative process of recreating something that has already been developed, whereas a development contract is creating something entirely new.

Contract Lengths

The fourth comparison analyzes by contract length. The Kruskal-Wallis test first determines if there is a significant difference between the short, medium, and long contracts. Then a Mann-Whitney test discovers which length of contract is different from the others. A short contract is defined as any contract that is 4 years or less, medium is 4 to 7 years, and long is over 7 years. The Kruskal-Wallis tests conclude significant differences of the CPI ranges at 35, 40, 45, 50, and 55% complete. Table 11 displays the results of the Mann-Whitney tests.

TABLE 11
Comparing CPI Ranges by Contract Length

Mann-Whitney test results for Contract Length																
Percent Complete	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Medium vs Short	-	-	-	-	-	R	R	R	R	R	-	-	-	-	-	-

Based on the results, there is a difference between the CPI ranges of the Medium contracts and Short contracts when they are 35% to 50% complete. For each of these instances, the short contracts have smaller medians, indicating they are more stable. There is only one instance where the CPI intervals possess a significant difference. There are no SPI(t) statistically significant results.

Platforms

Lastly, military platforms are compared with the Kruskal-Wallis test at an alpha of 0.05. The seven platform categories are: aircraft systems (AS), electronic/automated software systems (EAS), missile systems (MS), ordnance systems (OS), ship systems (ShS), space systems (SpS), and surface vehicle systems (SVS). Differences were not found for either CPI or SPI(t) between the various platforms.

Limitations

There are several limitations in this research. First, Over-Target-Baselines (OTBs) contracts removed from the dataset total to about 20% of all the available contracts. While removal of OTBs is consistent with previous DoD stability research, this is a large amount of data that was not able to be utilized. Second, the dataset contains contracts from only DoD ACAT I programs. The specific results should not be generalized to programs outside the DoD without confirmatory research being conducted within a program portfolio of interest and may not even be properly generalized to non-ACAT I programs in the DoD.

CONCLUSION

Some researchers and authors have taken the original Christensen and Heise research (1993) and indicated that their results were "generalizable," though the authors never made this claim. The concept of generalizability, or a rule of thumb, is an empirical matter. No claim is made in this paper either way. It is left to those EVM practitioners in the field to determine the applicability of the stability findings in this paper to analyses of their specific program(s). We recognize, however, that EVM usage is wide-spread across many countries and industries, encompassing both the public and private sector. It is not DoD unique. Thus, the results presented here provide general stability insights for all entities that employ EVM analysis in their procurement decisions. However, as with any dataset analysis, the *specific* results should not be extrapolated outside DoD's MDAP contracts which have not undergone an OTB. Importantly, according to Kristine Thickstun, there is no way to predict if a contract will have an OTB (Thickstun, 2010). Therefore, although this research shows certain stability characteristics when the contract *does not* have an OTB, the question of whether the contract will have an OTB or not remains. Thus, the applicability of stability properties to the contract remains tied to the unresolved question of being able to predict whether the contract will be OTB or not. This limitation is true of all the past DoD research as well since they too removed OTBs from the analysis.

The definition of "stability" has (understandably) morphed over time. *To answer the question of stability, then, is intricately tied to the definition used.* This research finds that CPI stability utilizing the *range* definition behave similar to past research and the "stability rule" but CPI stability utilizing the *interval* definitions stabilize later than the original "stability rule" states. SPI(t) behaves very similar to CPI when using the *range* definition of stability but stabilizes later in a contracts life using the *interval* definitions. From comparing CPI and SPI(t) stabilities among services, SPI(t) stabilities are very similar for all Services, while Army contracts' CPIs are either the same or less stable than AF and Navy's. When comparing contract types there are no differences in CPI stabilities, but the SPI(t) tends to be more stable in CP contracts. Between life-cycle phases, SPI(t) stability is very similar, but production contracts are more stable than development contracts in terms of CPI, using the range definition of stability (CPI intervals have no difference). The CPI stability ranges of contracts with different lengths vary, but only in the middle part of a contract. Comparisons between platforms show that the different platforms have no difference in CPI and SPI(t) stabilities.

The findings of this paper emphasize that the definition of stability used for the analysis matters. The obvious question is "which one is better"? The different definitions of stability have their advantages and disadvantages. The range definition is less dependent on a specific percent complete since it uses the maximums and minimums, whereas the interval definitions are more reliant on a single percent complete since that single point determines stability or not. The range definition takes into account the entire contract's life after a specific percent complete, but the absolute interval looks at a single point and compares it to the final. The range definition, however, is a little more complicated to comprehend and apply especially when using to prospectively predict contract performance. One does not know if the current CPI is the maximum, minimum, or somewhere in the middle of the contract's entire performance, so it is difficult to use a definition that utilizes the maximum and minimum. The interval definitions are easier to apply and more conservative, ultimately predicting a range of plus or minus 0.10 around the given CPI. The relative interval definition is simply another interpretation of the absolute interval definition. Therefore, if pressed to choose a single definition for stability to use in the future, we recommend the absolute interval definition. It is more dependent on a single percent complete, but it is easier to understand and more conservative. These two characteristics are important to program offices as they examine the performance of contracts and make future procurement decisions.

The implications of utilizing the absolute interval definition are clear. Stability, with this definition, occurs much later in the contract than previous research indicates. This delineates the window of opportunity where procurement officials can make positive changes with respect to cost and schedule performance.

Future research on stability is needed. Unanswered are the questions of the root causes of stability and an examination of the shocks that cause stability to waiver. This can be accomplished through an improved understanding of low-level project performance characteristics and their relationship to CPI and SPI(t).

REFERENCES

Abba, W. (2008). "The Trouble with Earned Schedule." *The Measurable News*, *36*(4): 28–30.

Antvik, S. (2001). "Why was Earned Value Management so Important to the Swedish Government in the Gripen Project?" Project Management Institute Contract Performance Management Conference, May 22-24, San Diego, California.

Bhosekar, S., & Vyas, G. (2012). "Cost Controlling Using Earned Value Analysis in Construction Industries." *International Journal of Engineering and Innovative Technology*, *1*(4): 324-332.

Christensen, D. (1996). "Project Advocacy and the Estimate At Completion Problem." *Journal of Cost Analysis*, *13*(1): 35–60.

Christensen, D. (1999). "Using the Earned Value Cost Management Report to Evaluate the Contractor's Estimate at Completion." *Acquisition Review Quarterly*, *19*: 283–296.

Christensen, D., & Payne, K. (1992). "Cost Performance Index Stability- Fact or Fiction?" *Journal of Parametrics*, *12*(1): 27–40.

Christensen, D., & Heise, S. (1993). "Cost Performance Index Stability." *National Contract Management Journal*, *25*: 7–15.

Christensen, D., & Templin, C. (2002). "EAC Evaluation Methods: Do They Still Work?" *Acquisition Review Quarterly*, 9: 105–116.

Crumrine, K., & Ritschel, J. (2013). "A Comparison of Earned Value Management and Earned Schedule as Schedule Predictors on DoD ACAT I Programs." *The Measurable News*, *41*(2): 37-44.

Defense Acquisition University (DAU). (2013). *Defense Acquisition Guidebook,* [On-line]. Available at https://acc.dau.mil/CommunityBrowser.aspx?id=488728 [Retrieved December 5, 2014].

Fleming, Q., & Koppelman, J. (1998). "Earned Value Project Management: A Powerful Tool for Software Projects." *Cross Talk, the Journal of Defense Software Engineering*, *11* (6): 19–23.

Fleming, Q., & Koppelman, J. (2008). "The Two Most Useful Earned Value Metrics: the CPI and the TCPI." *Cross Talk, the Journal of Defense Software Engineering*, 21 (12): 16–18.

Government Accountability Office. (2009). "*GAO Cost Estimating and Assessment Guide: Best Practices for Developing and Managing Capital Program Costs*." (GAO-09-3SP). Washington, DC: DIANE Publishing.

Government Accountability Office. (2013). "*Defense Acquisitions: Assessments of Selected Weapon Programs*." (GAO-13-294SP). Washington, DC: GAO.

Hayter, A. (2007). *Probability and Statistics for Engineers and Scientists* (3rd ed.). Belmont, CA: Duxbury.

Henderson, K., & Zwikael, O. (2008). "Does Project Performance Stability Exist? A Re-examination of CPI and Evaluation of SPI(t) Stability." *Cross Talk, the Journal of Defense Software Engineering*, *21*(4): 7–13.

Kutner, M., Nachtsheim, C., & Neter, J. (2004). *Applied Linear Regression Models*. New York: McGraw-Hill.

Lipke, W. (2003). "Schedule is Different." *The Measurable News*, *31*(4): 31-34.

Lipke, W. (2005). "Re-Examination of Project Outcome Prediction: Using Earned Value Management Methods." *The Measureable News*, *33*(2): 14–26.

Marshall, R. (2008). *The Top 10 Secrets of Earned Value Management*. [On-line]. Available at www.govtech.com/pcio/The-Top-10-Secrets-of-Earned.html [Retrieved December 5, 2014].

Popp, M. (1996). "Probability Distributions of CPI at Complete vs. CPI Today." Internal NAVAIR Report, Unpublished. Pax river, MD: US Navy Air Systems Command.

Singer, N. (1982). "Cost Growth in Weapon Systems: Recent Experience and Possible Remedies." Washington, DC: Congressional Budget Office (CBO), [On-line]. Available at http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/51xx/doc5102/doc19-entire.pdf [Retrieved October 11, 2014].

Society of Cost Estimating and Analysis (SCEA). (2010). "Module 15 Earned Value Management (EVM)." *Cost Estimating Body of Knowledge* (CEBoK). Ft. Belvoir, VA: Author.

Thickstun, K., & White, E. (2010). "Predicting Over Target Baselines (OTBs)." *The Measurable News*, *38*(4): 10–14.

Vandevoorde, S., & Vanhoucke, M. (2006). "A Comparison of Different Project Duration Forecasting Methods Using Earned Value Metrics." *International Journal of Project Management*, *24*(4): 289-302.