3-23-2018

An Analysis of the Estimate at Complete for Department of Defense Contracts

Deborah B. Kim

An Analysis of the Estimate at Complete for Department of Defense Contracts

THESIS

Deborah B. Kim, First Lieutenant, USAF

AFIT-ENC-MS-18-M-214

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENC-MS-18-M-214

An Analysis of the Estimate at Complete for Department of Defense Contracts

THESIS

Presented to the Faculty

Department of Mathematics and Statistics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Cost Analysis

Deborah B. Kim, BS

First Lieutenant, USAF

March 2018

AFIT-ENC-MS-18-M-214

An Analysis of the Estimate at Complete for Department of Defense Contracts

Deborah B. Kim, BS

First Lieutenant, USAF

Committee Membership:

Edward D. White, PhD
Chair

Jonathan D. Ritschel, PhD
Member

Chad Millette
Member

AFIT-ENC-MS-18-M-214

Abstract

When contractors provide timely, reliable, and actionable information on the status of a contract, both contractors and government program offices can provide an accurate estimate of a contract's completion cost to leaders who can then take proactive course corrections if there are emerging problems in a program. This research shows that the cumulative cost performance indices provided by contractors and program offices are high and less accurate than those of previous years and/or that a significant amount of ACWP is being documented in the final portion of a contract. This research replicates Christensen's findings in 1996 which proved that using the SCI to calculate EAC (EAC_{SCI}) was a more accurate indicator of the final cost vice the CPI (EAC_{CPI}). Christensen's research showed that using EAC_{SCI} to predict the final cost resulted in a deviation of only 5% starting at the 20% complete point whereas EAC_{CPI} took until the 70% complete point. Consistent with Christensen's research, EAC_{SCI} is still a more accurate indicator of CAC according to this study but not by a significant amount. When EAC_{SCI} is used to predict the final cost on contracts from the 21st century, a 5% deviation from the final cost starts at the 70% complete point. This research shows that data integrity has suffered since Christensen's research in 1996 and that there is no significant difference between using CPI or SCI as the performance indicator to predict the final cost at complete.

Acknowledgments

I would like to thank my family, classmates, golf friends, and yoga sisters for all the good memories I've had during my time at AFIT.

I would also like to thank my research advisor, Dr. Edward White, for pushing me past my comfort zone which therefore helped me grow as an individual during my time at AFIT. I also wish to express the deepest gratitude towards Mr. Chad Millette for going beyond his means to answer my questions while helping me learn and understand my research.

Deborah B. Kim

Table of Contents

List of Figures

List of Tables

Page

I. Introduction

General Issue

The Estimate at Completion (EAC) is an Earned Value Management (EVM) metric used to predict the final cost of a program. Often, the EACs calculated by the government and contractors are significantly lower than what ends up being the final contract cost. Furthermore, pressure from stakeholders to keep programs from being cancelled has led to a toxic culture of reporting optimistically low EAC calculations (Christensen, 1996). Organizations encourage goals that are unrealistic because there is an over optimism that these goals are attainable. This mentality has resulted in programs that cost more than planned and produce results that do not satisfy all requirements (GAO, 2009).

Two performance indices used for the EAC are the Cost Performance Index (CPI) and Composite Index (SCI) (DAU, 2017). Based on Christensen's research (1996), the final cost of a program is quickly and accurately predicted when SCI is used as the performance index when calculating EAC (EAC_{SCI}). SCI is the product of the CPI times the Schedule Performance Index (SPI). Christensen's comparisons showed that the cost overruns projected by the contractor and government were unreasonably optimistic throughout the lives of the contracts examined (1996). Contractors strategically propose low cost estimates, and when a program's budget is based on these low cost estimates it becomes apparent that either the developer or customer must pay for the resulting cost overrun (GAO, 2009).

Problem Statement

The purpose of this research is to determine if EAC_{SCI} is still a more accurate predictor of the final Cost at Completion (CAC) over EAC_{CPI} for current contracts in the Department of Defense (DoD). Christensen used data from 64 contracts and found that EAC_{SCI} predicts the CAC more accurately than EAC_{CPI} (1996). Christensen's work has not been updated since 1996 and has become routinely cited by subsequent EVM authors of academic literature. In practice, System Program Offices (SPOs) use multiple methods to calculate EAC, to include CPI, SPI, SCI, and weighted indices. By using data from current DoD contracts, this research will determine whether using SCI to calculate EAC is still a more accurate method of predicting CAC.

Participants involved in the procurement of acquisition programs have learned how to routinely calculate minimum and maximum EACs using CPI and SCI, respectively. According to Christensen's analysis, estimating the final CAC using EAC_{SCI} is a quicker and more accurate predictor of CAC versus using EAC_{CPI}. When Christensen used SCI as the performance indicator to calculate EAC (1996), there was less than a 5% deviation from the final cost at a contract's 20% complete point (see Figure 1). However, when Christensen used CPI as the performance indicator to calculate EAC, it was not until the 70% complete point that there was a less than 5% deviation from the final CAC. Using SCI as the performance indicator was a quicker method of calculating the final CAC according to his research.

Currently, the DAU's (Defense Acquisition University) gold card shows that EAC is calculated either by using CPI or the SCI. If CPI is still the most optimistic

method of calculating the EAC (Christensen, 1996), then program managers should be cautious in using this metric as an estimating tool. Moreover, if SCI is still the most accurate measurement tool in predicting the final cost of a contract, SCI should be used in lieu of other performance indices.

Research Questions

The goal of this research is to identify which performance index most accurately predicts the final CAC. Accuracy is determined by calculating the deviation from the cost at complete (%DCAC) using performance indices CPI and SCI. An accurate estimate is defined as being within 5% of the final CAC for this research.

1. Which efficiency factor is most accurate at predicting CAC?
2. At what percent complete does the EAC get within 5% of the CAC?
3. What are the major and moderate drivers that influence |%DCAC|?

Methodology

EVM data pulled by the Cost Assessment Data Enterprise (CADE) portal on 26 July 2017 was used to calculate and compare EACs for this research. This dataset consisted of 167 programs, 451 contracts, and 863 contract line item numbers (CLINs) from years 2000 through 2017. CADE is an Office of the Secretary of Defense Cost Assessment and Program Evaluation (OSD CAPE) initiative created to increase the effectiveness of displaying data on a single web-based application to improve reporting compliance and source data transparency. This research used only completed programs to predict EAC and therefore no programs initiated in 2017 were included in this study.

There were three parts to this research. The first part was to replicate Christensen's study from 1996 to determine whether EAC_{SCI} is still the quickest method of predicting CAC. Figure 1 was recreated using the data provided by CADE. Consistent with Christensen's research (1996), the CLINs used to recreate this graph had no over target baselines (OTB), started reporting at ≤ 20%, and reached 100% completion. Next, all CLINs that had no OTBs and reached ≥ 92.5% completion were used for the following two parts of the research. The 92.5% completion point signifies completion based on the research by Tracy and White (2011). The second part re-analyzed Tracy and White's work to examine contracts with respect to deviation from the final cost at complete (CAC) with hypothesis tests to include t-tests as well as Wilcoxon rank-sum tests. The research concludes with a population analysis to identify major and moderate drivers in predicting the |%DCAC| through Ordinary Least Squares (OLS).

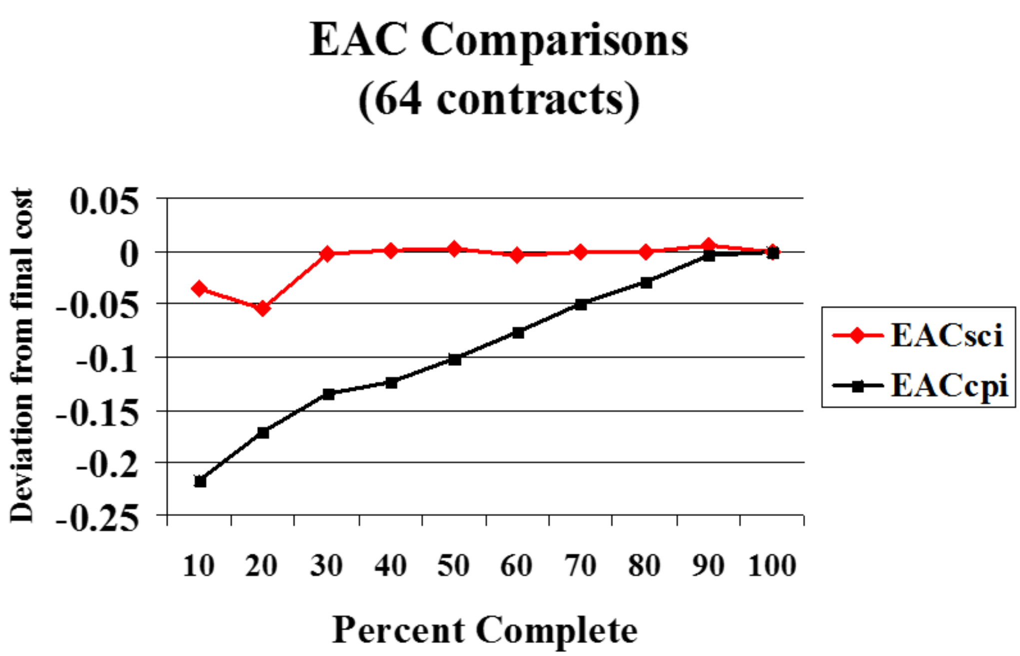


Figure 1. Christensen's EAC Comparisons (1996)

Assumptions/Limitations

The scope of this thesis covers all completed CLINs from years 2000 through 2017 with available EVM data from CADE. Because of changes in EVM reporting procedures during this timeframe, the quality of the data may have potentially been affected. In 2005, reporting requirements for Contract Performance Reports (CPR) changed from the original requirements in 2003. In 2012, the CPR changed to the Integrated Program Management Report (IPMR). This research assumes that the data provided by CADE during 2000-2017 is accurate, and subsequently that the metrics reported by contractors and program offices are accurate.

Thesis Overview

The next section of this research, Chapter 2, provides a literature review of EVM and background information on previous studies of EAC. Chapter 3 outlines the methodology as well as the analysis and results of the research. Chapter 4 is the journal article that was written for the *Journal of Public Procurement*, and Chapter 5 discusses the results of this research and potential ideas for future research.

II. Literature Review

Overview

This chapter reviews previous research conducted on calculating the EAC as well as the different definitions of stability. First, it discusses background information on the Earned Value Management System (EVMS) by defining key terms, presents a historical outline of Acquisition Reforms, and highlights its current use in the DoD. Next, it introduces EAC's role in EVMS and defines the four standard performance indices to predict the final cost of a program. This chapter then concludes with a discussion of previous research efforts related to EAC.

Earned Value Management

EVM is "an industry standard method of measuring a project's progress at any given point in time, forecasting its completion date and final cost, and analyzing variances in the schedule and budget as the project proceeds. It compares the planned amount of work with what has actually been completed, to determine if the cost, schedule and work accomplished are progressing in accordance with the plan" (Lessard & Lessard, 2007: 45). Government and contractor Program Managers (PMs) use EVM to track time, budget, and performance goals on programs and are responsible for the development, production, and sustainment of a program's objectives to meet the end user's operational needs (DoD).

The PM then reports the cost, schedule, and performance of a program to the Milestone Decision Authority (MDA) to determine if the program can enter the next phase of acquisition. Next, the MDA reports and updates the program's performance to

higher authority, including Congress, from reports collected by the PM (DoD). For this reason, EVM metrics gathered by the PM are crucial for the sustainment of a program's development because it provides a "joint situational awareness of program status" in cases where "proactive course corrections" are needed (Kranz & Bliss, 2015: 5). EVM metrics can show emerging problems in a program so that leaders can take corrective action that will limit the damage done to a program's cost and schedule goals (Department of the Air Force, 2007).

EVM is required for all cost or incentive contracts equal to or greater than $20 million and/or have a high risk in development work for the government (Department of the Air Force, 2007). EVM is best suited for projects that have defined deliverables, longer durations, strict budget limits, and a single contract encompassing all or most of the effort. EVM is less suited for projects that are difficult to define or have open-ended objectives, shorter durations, and use Level of Effort (LOE) support hours as the primary deliverable (Rose, 2002). Appendix A shows when EVM is required for contracts.

A version of EVM was first introduced to the DoD in the 1960s and was called the Cost/Schedule Control Systems Criteria (C/SCSC) approach (Department of the Air Force, 2007). In accordance with DoD 7000.2, the DoD set 35 C/SCSC as a standard for all programs in 1967. Later, in 1998, the DoD adopted the American National Standards Institute/Electronic Industries Alliance standard ANSI/EIA-748 for major defense acquisition programs. The implementation of ANSI/EIA-748 reduced the number of criteria from 35 to 32, but it is still a complex and heavily regulated governing approach with substantial bureaucracy and far too many non-value-added requirements. (See

Appendix B for a complete list of the 32 criteria.) Table 1 shows the requirements for each contract level separated by dollar amount.

Table 1. Requirements Based on Contract Dollar Amount (EVMIG, 2006)

<table>
<tr><th>≥$50 million</th><th>REQUIRED</th></tr>
<tr><td>Includes: Contracts for highly classified, foreign, and in-house programs.</td><td rowspan="4">• Must use ANSI/EIA-748 complaint and validated management system.
• CPR (all formats) is required.
• Integrated Master Schedule is required.
• Schedule Risk Assessment (SRA) is required.</td></tr>
<tr><td>Not required for: Firm-fixed price contracts. (Business case analysis and MDA approval required.)</td></tr>
<tr><td>Not recommended for: Contracts less than 12 months in duration.</td></tr>
<tr><td>May not be appropriated for: Non-schedule base contract efforts, e.g., level of effort.</td></tr>
</table>

<table>
<tr><th>≥$20 million but < $50 million</th><th>REQUIRED</th></tr>
<tr><td>Includes: Contracts for highly classified, foreign, and in-house programs.</td><td rowspan="2">• Must use ANSI/EIA-748 complaint management system. No validation.
• CPR Formats 1 and 5 are required.
• Integrated Master Schedule is required.</td></tr>
<tr><td>Not required for: Firm-fixed price contracts. (Business case analysis and MDA approval required.)</td></tr>
<tr><td>Not recommended for: Contracts less than 12 months in duration.</td><td>OPTIONAL</td></tr>
<tr><td>May not be appropriated for: Non-schedule base contract efforts, e.g., level of effort.</td><td>• CPR Formats 2, 3, and 4 are optional.
• Schedule Risk Assessment is optional.</td></tr>
</table>

<table>
<tr><th>< $20 million</th><th>OPTIONAL – USE JUDGMENT</th></tr>
<tr><td>Evaluate management needs carefully to ensure only minimum information needed for effective management control is requested.</td><td rowspan="4">• ANSI/EIA-748 compliance is discretionary and should be based on risk.
• CPR Formats 1 and 5 are recommended.
• Integrated Master Schedule is optional.</td></tr>
<tr><td>Requires cost-benefit analysis and PM approval.</td></tr>
<tr><td>Not recommended for: Contracts less than 12 months in duration.</td></tr>
<tr><td>May not be appropriated for: Non-schedule base contract efforts, e.g., level of effort.</td></tr>
</table>

Since the DoD began utilizing EVM, there have been many changes to reporting procedures. The Contract Performance Reports (CPRs) were used as the primary documenting method between contractors and PMs for contracts that required the use of EVM. The CPR provided the cost and schedule performance of a program early in the acquisition contract to forecast future contract performance. The CPR requirements were established in 2003 under DoD 5000.02, and the requirements were subsequently revised in 2005 (USD-AT&L, 2008). The revisions only applied to contracts after its release. Therefore, ongoing contracts awarded before 2005 did not have to adopt the changes released in 2005. All CPRs were collected by the Office of the Under Secretary of Defense (Acquisition, Technology, and Logistics) and are located in the Central Repository (CR). In 2007, the CR began collecting CPRs from before and after the revision.

The CPR's five formats are the Work Breakdown Structure (WBS), Organizational Categories, Baseline, Staffing, and Problem Areas. In 2012, the Defense Contract Management Agency (DCMA) established the IPMR which combined and updated the CPR and the Integrated Master Schedule (IMS) Data Item Descriptions (DID) (EVMS PAP, 2012). All contracts post July 2012 require IPMRs in lieu of CPRs. The IPMR is scheduled around seven formats that contain the content from CPRs with the addition of the electronic file of the contractor's Integrated Master Schedule (IMS) and the annual report in the contractor's electronic file format. Table 2 shows the correlating CPR DID & IMS DID to the IPMR DID. Because of the reoccurring changes

in reporting rules over the years, the quality of EVM data for this research has potentially been affected.

Table 2. Correlating CPR and IMS with IPMR DID (EVMS PAP, 2012)

CPR DID & IMS DID	IPMR DID
CPR Format 1	IPMR Format 1
CPR Format 2	IPMR Format 2
CPR Format 3	IPMR Format 3
CPR Format 4	IPMR Format 4
CPR Format 5	IPMR Format 5
IMS	IPMR Format 6
N/A	IPMR Format 7

EVM uses various reporting metrics to objectively quantify a program's performance. The primary components of EVM are the Budgeted Cost of Work Performed (BCWP), Actual Cost of Work Performed (ACWP), and Budgeted Cost of Work Scheduled (BCWS). These components are the foundations for Schedule Variance (SV), Cost Variance (CV), EAC, Total Allocated Budget (TAB), and Budget at Complete (BAC). Figure 2 shows the main elements of EVM from Defense Acquisition University (DAU).

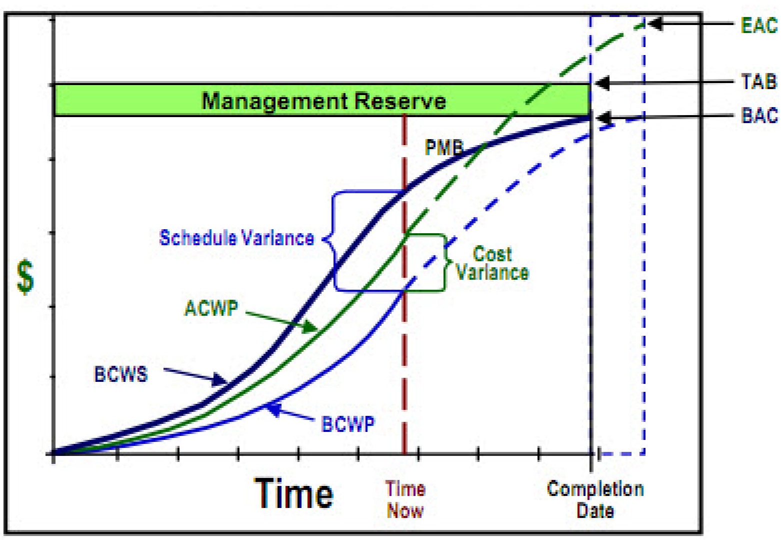


Figure 2. Main Elements of EVM (DAU, 2017)

BCWP, also known as Earned Value (EV), is the value of the work accomplished up to a point in time. Actual Cost (AC) is another term for ACWP and is the cumulative cost "spent to a given point in time to accomplish an activity, work package, or project and to earn the related value" (Anbari, 2003:13). BCWS, also known as Planned Value (PV), is the approved time-phased budget baseline "in which contract EVM performance is measured" (Department of the Air Force, 2007:23); BCWS and the Program Management Baseline (PMB) are used interchangeably. SV is the difference between BCWP and BCWS, and CV is the difference between BCWP and ACWP. The Total Allocated Budget (TAB) is the sum of all budgets for work on a contract to include the Management Reserve (MR). The vertical "Time Now" line in Figure 2 shows that this program is both behind schedule and over-budget since there is a negative SV and a negative CV. The EAC is the ACWP plus the estimated cost of the remaining work. Table 3 defines the main elements of EVM.

Table 3. Summary of EVM Measurements (DAU, 2017)

EVM Measurement	Meaning
BCWP	Value of work accomplished, also known as EV
ACWP	Cost of work accomplished, also known as AC
BCWS	Value of work planned to be accomplished, also known as PV
OTB	Sum of Contract Budget Bases (CBB) and recognized overrun
TAB	Sum of all budgets for work on contract
MR	Budget withheld by PM for unknowns/risk management
EAC	Estimate of total cost for total contract through any given level

Estimate at Completion

While EVM interprets historical data, EAC is an estimating tool used to calculate the final cost of a program. EAC is also known as the Latest Revised Estimate (LRE). Analysts typically compute EAC by using the "Cost Performance Index (CPI) or the

Schedule Performance Index (SPI) as products of the Earned Value Management System" (Tracy & White, 2011:191). The CPI is the ratio between BCWP and ACWP; a CPI greater than 1 means that a program is under-budget, a CPI equal to 1 means that the program's budget is on target, and a CPI less than 1 means that a program is over-budget. SPI is the ratio between BCWP and BCWS. Similar to CPI, a SPI greater than 1 means that a program is ahead of schedule, a SPI equal to 1 means that a program's schedule is on target, and an SPI less than 1 means that a program is behind schedule. The typical condition of a defense contract is over-budget and behind schedule and therefore the CPI and SPI are usually below 1 (Christensen, 1996).

Additionally, one of the main shortfalls of determining the current state of a program's schedule is that SPI tends to 1 and SV tends to 0 at the end of a program regardless of performance. The variables used to solve for SPI and SV are BCWP and BCWS which are both measured in units of dollars. When a program is completed, all of the planned work, BCWS, equals all the performed work, BCWP. Therefore, SPI becomes 1.0 and SV becomes 0 at the end of a program. Lipke introduced Earned Schedule (ES) as an extension of EVM (2003). Instead of using costs to predict schedule, ES uses measurements of time to calculate SPI and SV. When ES is used to solve for SPI and SV, they are noted SPI(t) and SV(t). Unlike EVM, ES allows a program to have a SPI(t) less than 1 and a negative SV(t) at the end of a contract's life. ES more accurately shows how a program performed at the end of its lifetime than EVM.

EAC is calculated during the progression of a contract's development, and the goal is to keep costs in line with the original TAB. Although there are multiple ways to

calculate EAC, the most commonly used techniques are to use the CPI and SCI as performance indicators. Generally, the most optimistic EAC occurs when CPI is used as the performance index in Equation 1 to calculate CAC and the most pessimistic estimate occurs when the SCI is used in Equation 1 to calculate CAC. An EAC is called optimistic because it is potentially the lowest cost that a program could cost with all other factors remaining constant with the original cost schedule. Christensen used the following equations to calculate the EACs of contracts. Equations 2 and 3 are used to estimate the floor (minimum) and ceiling (maximum), respectively.

$$\mathbf{EAC = Actuals\ to\ Date + [(Remaining\ Work)/(Performance\ Factor)]} \quad (1)$$

$$\mathbf{EAC_{CPI} = ACWP_{CUM} + [(Final\ BAC - BCWP_{CUM})/CPI_{CUM}]} \quad (2)$$

$$\mathbf{EAC_{SCI} = ACWP_{CUM} + [(Final\ BAC - BCWP_{CUM})/(CPI_{CUM} * SPI_{CUM})]} \quad (3)$$

In addition to CPI and SCI, EAC can also be calculated by using a weighted performance index. CPI and SPI are assigned weights W1 and W2, and these weights must sum to one. Table 4 shows the breakdown of the four main types of performance indices used to calculate EAC.

Table 4. Summary of Performance Indices

Performance Index	Formula
CPI	BCWP / ACWP
SPI	BCWP / BCWS
Weighted Index	W1*SPI + W2*CPI
SCI	SPI*CPI

There are many ways to predict CAC to include the four performance indices shown in Table 4. Christensen analyzed 64 completed contracts from the Defense

Acquisition Executive Summary (DAES) to determine which performance index most closely predicted the final CAC (1996). His research concluded that the EAC based on CPI was a reasonable lower bound to the final cost of a defense contract and that "estimates supported by government and contractor management were not significantly different from the CPI-based EAC" (Christensen, 1996:7). EAC was most accurately measured when SCI was used as the performance index but contractor and government program managers tend to report the lower estimates provided by the CPI performance index. Christensen's study (1996) is shown in Figure 1 and shows that EAC_{CPI} deviates from the final cost significantly, especially when compared to EAC_{SCI}. EAC_{SCI} deviates less than 5% at the 20% complete point whereas EAC_{CPI} deviates less than 5% at the 70% complete point. According to Christensen's analysis, the EAC_{CPI} deviates a great deal and is inaccurate until the contract is near completion. The objective of this study is to replicate Figure 1 using contracts from 2000-2017.

Christensen defined an overrun as being the difference between the cumulative BCWP and the cumulative ACWP. Based on his analysis of the 64 completed contracts, his comparisons showed that the overruns projected by the contractor and government were "excessively optimistic throughout the lives of the contract examined" (Christensen, 1994:25). Figure 3 shows the over optimism in estimates by both contractors and the government.

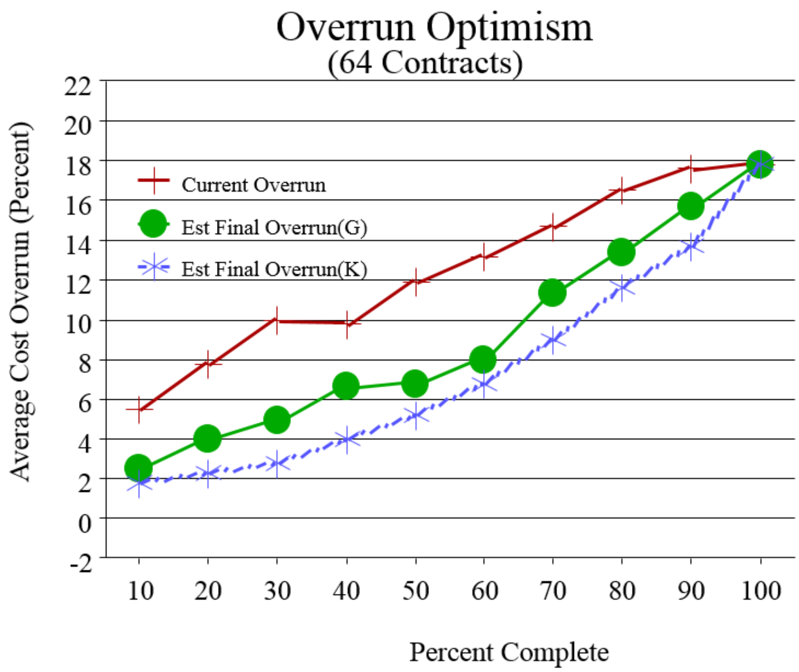


Figure 3. Overrun Optimism (Christensen, 1992)

CPI's Stability Rule

Although there is currently no research on SCI stability, there are many different definitions of CPI stability. Christensen and Payne (1992) established that the cumulative CPI on completed Air Force contracts did not change by more than 10 percent from the value at the 20 percent contract completion point. This assertion is pertinent to this thesis because CPI is one of the performance indices used to calculate EAC. Equation 2 uses the cumulative CPI of a contract to calculate the minimum EAC. The To Complete Performance Index (TCPI) represents the CPI in which the contractor must perform in the remaining work to meet the budgetary goal. If the TCPI is much higher than the current CPI, then the contractor will have to either significantly improve the efficiency of the remainder of the program's budget or the contract will be over budget.

Re-evaluating CPI's Stability Rule and Evaluating SPI(t)'s Stability

Christensen's stability rule was re-evaluated by Petter, Ritschel, and White (2015) by using current data to explicitly state the multiple definitions of the "stability rule" in EVM literature as well as examining its effects on SPI(t); they defined the classifications for range, absolute interval, and relative interval (see Table 5) based on past researchers. Petter et al.'s research re-examined the existence of the CPI stability rule to determine the percent complete point where stability is achieved. Because of the limited amount of research on Earned Schedule's SPI(t) stability, Petter et al. applied the same process of determining the stability of CPI to SPI(t). Earned Schedule (ES), unlike EVM, uses time to measure schedule; SPI(t) means that time is used as the unit of measurement for schedule. Then, they compared different categories of contracts to determine if stability properties varied by category.

Four different comparisons were made for the comparison analysis by category. Table 6 shows the categories for comparison analysis. This research found that the range definition of stability for CPI is consistent with past research. However, the absolute interval stabilizes, at the earliest, during the 45 percent complete point, and the relative interval stabilizes, at the earliest, during the 50 percent complete point. SPI(t) performs similarly to CPI when using the range definition of stability, but SPI(t) stabilizes later in a contract's life for the absolute interval and relative interval at 50 percent and 65 percent, respectively. SPI(t) is similar among all services, but the Army's CPIs are either the same or less stable than those of Air Force and Navy. For contract types, there are no differences in CPI stability but SPI(t) tends to stabilize more in Cost Plus contract. Using

the range definition of stability, the life-cycle phases are similar in terms of SPI(t) but production contracts are more stable in terms of CPI. There are no significant differences between different military platforms for CPI ranges and SPI(t) ranges and intervals.

Table 5. Stability Definitions (Petter, Ritschel, & White, 2015:348)

Definition Name	Stability Definition	Stability 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)
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)

Table 6. Categories for Comparison Analysis (Petter, et.al, 2015:352)

Categories			
Services	**Contract Types**	**Life-cycle Phases**	**Platforms**
Air Force	Fixed Price	Development	Aircraft System
Army	Cost Plus	Production	Electronic/Automated System
Navy			Missile System
			Ordnance System
			Ship System
			Space System
			Surface Vehicle System

Calculating EAC with Multiple Regression

Time series forecasting techniques, linear and non-linear regression based analyses, Bayesian probability, and other methods have been used to calculate EAC. The most recent use of predicting CAC using multiple regression was studied by Tracy and White. Tracy and White state that "accurate EACs are those that most closely estimate the final cost of the contract" (Tracy & White, 2011:193) which they deem the CAC. In lieu of using the four main performance indices (CPI, SPI, SCI, and the weighted index) to calculate EAC, Tracy and White's research provided five working multiple regression models to accurately predict the final cost of the average major weapons system contract using Contract Performance Report (CPR) data. Tracy and White identified the 92.5% completion point of a contract to have no statistical difference from the 100% completion point and therefore used all contracts that were completed up to the 92.5% complete for the regression models. Tracy's final data comprised of 51 programs, 241 contracts, and 3,725 reports from 5 Navy programs and 46 Air Force programs from the DAES database.

The Contractor Estimate at Completion (CEAC) was the main driver for three of the five models (at the 35, 50, and 65 percent complete points), and the two of the other models (at the 25 and 75 percent complete points) used the Budget at Completion (BAC) as the main driver. In their study, Tracy and White included contracts with an Over Target Baseline adjustment.

Summary

This chapter reviewed previous work conducted on EAC and the most current changes to EVM reporting procedures. The quick rule of thumb to estimate minimum and maximum EAC has been to use CPI and SCI as performance indices. Instead of automatically calculating a lower estimate using CPI and an upper estimate using SCI, this thesis seeks to find whether SCI is still the most timely and accurate method of predicting a reliable EAC for current DoD contracts.

EACs are crucial for the sustainability of a contract's life, and it is essential that both government and program offices accurately predict CACs to budget properly for DoD programs. Christensen laid the foundation for future research on EAC's properties. Petter, Ritschel, and White re-evaluated Christensen's CPI stability rule and assessed SPI(t)'s stability properties. Tracy and White created five multiple regression models to predict EACs based on the percent complete of a program. The next chapter seeks to take the literature reviewed here to analyze 21st century data on which performance index is the most accurate predictor of EAC for timeliness and accuracy, re-analyze a portion of Tracy and White's work, and then use a population analysis to identify main drivers of deviations from final CAC.

III. Methodology/Analysis & Results

Overview

Chapter 3 serves as a dual purpose to both outline the steps taken for this research while explaining the results along the process. This section will provide the source of the data, the limitations of EVM that were discovered along the way, the data included and excluded for the research, the formulas used for the study, and the statistical process used to perform the analyses. There are three parts to the analysis: the first analysis replicates Christensen's paper on the "Project Advocacy and the Estimate at Completion Problem," the second analysis focuses on re-examining a part of Tracy and White's "Estimating the Final Cost of a DoD Acquisition Contract," and the third analysis uses a population analysis to determine major and moderate drivers of $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$.

Data

The dataset used for this research is from CADE. Data on the cost and schedule of programs on major defense contracts are prepared by the contractor and program manager and sent to CADE through IPMRs. The mission of CAPE is to provide independent program analyses requested by the Under Secretary of Defense for Acquisition, Technology and Logistics (USD(AT&L)) and Congress. The key collaborators for CADE are the Program Offices, Service Cost Centers, AT&L PARCA and CAPE to comprise data for all DoD acquisition programs.

CADE is the central repository for all ACAT I EVM data and consists of Contractor Cost Data Reports (CCDRs/1921s), IPMRs, Cost Analysis Requirements Descriptions (CARDs)/ Technical Data (1921-Ts), Software Resource Data Reports

(SRDRs), Institutional Knowledge, and Policy Updates and Changes. IPMRs are submitted by the program offices and have monthly and quarterly updates for all the Work Breakdown Structures (WBS) for a project. The data provided by CADE matched the data from the IPMRs provided by each of the contract's respective program offices.

For this research, the data was provided directly by CADE. To calculate EAC from current data, the DoD contracts used for this study were from between the years of 2000 and 2017. Since the data was provided in real time, some of the contracts provided by CADE had not yet reached completion. Moreover, some of the contracts did not start reporting metrics until much later in the contract's lifetime. The BCWS, BCWP, and ACWP provided in the dataset are cumulative numbers. In the original dataset, there were 167 programs, 451 contracts, and 863 CLINs.

Calculations

The following calculations were performed on the original 863 CLINs before the data cleaning process. For each data point, the four EVM measurements used for the calculations were the BCWP, BCWS, ACWP, and the BAC. Explanations for these four variables can be found in Chapter 2.

Calculating Percent Complete

The formula used to calculate percent complete for this study was Cumulative BCWP/Final BAC. Christensen used this same formula to calculate percent complete for his analysis in 1996. He used other bases such as the contract budget base and the total allocated budget, but the results were insensitive for the choice of base. The BACs often change throughout the life of a contract. If the BAC increases a substantial amount from

month to month, the percent complete could decrease if the BAC is larger than the cumulative BCWP. This leads to inconsistent results and therefore it was determined that using the last reported BAC for each CLIN would be used for calculating percent complete. This allows for a stable denominator to ensure that the percent complete moves in one direction throughout the life of a CLIN.

The analysis for Part I required the percent completes to be bucketed in 5 percent increments from 0% to 100%. Most of the percent complete calculations were not in exact 5 percent increments such as 15% and 20%. Instead, the calculations resulted in decimals such as 15.896% or 20.112%. Identical to Tracy and White's methodology in defining percent complete, any percent complete within 2.5% of a specific percent complete bucket was determined to be that specific percent complete bucket. For example, if a certain CLIN was 6.8% complete, it would round down to 5% and if it also had a 4.36% complete point it would round up to 5%.

Calculating CPI and SCI

For every time a CLIN's metrics were reported, the CPI and SCI were calculated. CPI is calculated using $BCWP_{CUM}/ACWP_{CUM}$. To calculate SCI, SPI was first calculated using the formula $BCWP_{CUM}/BCWS_{CUM}$. Multiplying SPI and CPI results in the calculation for SCI. After calculating CPI, SPI, and SCI, it appears that many of these numbers were above 0.9, indicating that the contracts were performing well in terms of budget and schedule. 87.10% of the reported CPIs were above 0.9, 86.52% of SPIs were above 0.9, and 76% of SCIs were above 0.9.

After interviewing experts in the field of cost analysis, it became apparent that many did not believe in the confidence of EVM reporting. Mr. Wayne Abba, a former program analyst for contract performance management in the Office of the Under Secretary of Defense (Acquisition &Technology), agreed that the CPIs and SPIs reported by contractors and program managers are idealistic and inflated. He stated that contractors have learned how to manage data in the face of this misguided policy, thereby crippling its utility to actually track progress and inform management decisions.

Calculating EAC_{CPI} and EAC_{SCI}

After the CPIs and SCIs were calculated, the next step was to calculate EAC_{CPI} and EAC_{SCI}. The EAC_{CPI} was calculated using Equation 2 and EAC_{SCI} was calculated using Equation 3. The ACWPs, BCWPs, CPIs, and SPIs were cumulative values. However, the BAC value used to calculate EAC_{CPI} and EAC_{SCI} was the final reported BAC for each CLIN. Then, it was determined that using the last reported BAC for each CLIN should be used for Equations 2 and 3 because it is the most up to date BAC.

Calculating $\%DCAC_{CPI}$ and $\%DCAC_{SCI}$

Once the EACs were calculated, the deviations were calculated using Equations 4 and 5. Unlike Equations 2 and 3, Equations 4 and 5 used the Final ACWP number versus the $ACWP_{CUM}$ number. The Final ACWP value was the highest reported ACWP for each CLIN. It was determined that using the highest reported ACWP for each CLIN should be used instead of the last reported ACWP. Sometimes, the last reported ACWP's value was lower than the ACWP at an earlier point of a CLIN. Since ACWP shows the actual cost at a certain point in time for a CLIN, using the highest ACWP should be used for

Equations 4 and 5. After solving for the $DCAC_{CPI}$ and $DCAC_{SCI}$ using Equations 4 and 5, these values were multiplied by 100 to get these values in percentages.

$$\mathbf{DCAC_{CPI} = (EAC_{CPI} - Final\ ACWP) / Final\ ACWP} \quad (4)$$

$$\mathbf{DCAC_{SCI} = (EAC_{SCI} - Final\ ACWP) / Final\ ACWP} \quad (5)$$

Percent Complete Buckets and Averages

After calculating $\%DCAC_{CPI}$ and $\%DCAC_{SCI}$ for each CLIN, the $\%DCAC_{CPI}$ and $\%DCAC_{SCI}$ were averaged if there was more than one point that was within a certain percent complete bucket. For example, in CLIN "W58RGZ-09-C-0151 PRTA," the three percent completes that rounded to the 10% bucket were 8.521%, 9.973%, and 11.455%. The CPI, SCI, EAC_{CPI}, EAC_{SCI}, $\%DCAC_{CPI}$, and $\%DCAC_{SCI}$ were calculated for each of these three percent completes and are shown in Table 7. Since these three percent completes rounded to the 10% bucket, the average $\%DCAC_{CPI}$ and $\%DCAC_{SCI}$ were calculated to determine the average $\%DCAC_{CPI}$ at 10% and average $\%DCAC_{SCI}$ at 10%. This method was used for all CLINs that had multiple percent completes within a certain percent complete bucket.

Table 7. Calculating Average %DCACs for 10% Bucket

Actual %	CPI	SCI	EAC_{CPI}	EAC_{SCI}	$\%DCAC_{CPI}$	$\%DCAC_{SCI}$	Avg $\%DCAC_{CPI}$	Avg $\%DCAC_{SCI}$
8.52	1.17	1.23	27,000,937	26,000,937	-58.51	-60.06	-56.60	-54.28
9.97	1.08	1.10	29,089,863	29,089,963	-54.84	-55.31	-56.60	-54.28
11.45	1.12	0.87	34,192,877	34,192,877	-56.44	-47.47	-56.60	-54.28

Data Cleaning

There were two datasets used for this research. The first dataset comprised of 96 CLINs that represented the same characteristics as those of Christensen's 64 contracts (1996). The second dataset was broader in scope and consisted of 254 CLINs. The first dataset was used to replicate Christensen's study while the second dataset aimed to re-examine Tracy and White's study (2011) in terms of |%DCAC| and then determine major and moderate drivers of |%DCAC| using a population analysis.

Table 8 shows the data cleaning process for Part I of the analysis starting with the original 863 CLINs and 451 contracts. There were many CLINs for each contract, and although this research focuses on the CLINs, the number of contracts affected were still documented in the process. 145 CLINs had OTBs which reduced the number of CLINs to 718. If a CLIN did not report up to the 99.5% complete point (representative of 100% complete), it was removed from the dataset; a total of 590 CLINs were affected by this criterion which left a total of 128 CLINs. Then, if a CLIN did not start reporting until after the 20% complete point, it was removed from the dataset which reduced the dataset to 113 CLINs. Lastly, if a CLIN's $\%DCAC_{CPI}$ and $\%DCAC_{SCI}$ did not reach within 5% by the end of its reporting period, it was removed from the dataset. This left a total of 96 CLINs for Part I of the analysis.

Table 8. Dataset Characteristics for Part I Analysis

Criteria	Affected Contracts	Affected CLINs	Total Contracts	Total CLINs
Initial data extraction from CADE's website	451	863	451	863
Experienced an Over Target Baseline	67	145	384	718
Did not report to at least 99.5% complete	297	590	87	128
Does not meet percent complete requirement of ≤ 20%	10	15	77	113
EAC failed to reach within 5% of the final BAC	10	17	67	96

The second dataset used for Parts II and III of the analysis is shown in Table 9. The data cleaning process for the second dataset started off identical to the first dataset by removing CLINs that had OTBs. To re-examine Tracy and White's research regarding 92.5% as equivalent to 100%, CLINs that did not report to at least 92.5% were removed which left a total of 370 CLINs. Moreover, if a CLIN's $\%DCAC_{CPI}$ and $\%DCAC_{SCI}$ did not reach within 10% by the end of its reporting period, it was removed from the dataset; this left a total of 318 CLINs. Then, if a CLIN's $\%DCAC_{CPI}$ and $\%DCAC_{SCI}$ did not reach within 5% by the end of its reporting period, it was removed from the dataset; this left a total of 273 CLINs. Lastly, if a CLIN's $\%DCAC_{CPI}$ and $\%DCAC_{SCI}$ reached within 5% but occurred after 100% complete, it was removed from the dataset. Of the 254 CLINs, 147 completed at or greater than 92.5% but less than 99.5% while the remaining 107 reported a completion percentage of 99.5% or greater.

Table 9. Dataset Characteristics for Parts II and III of Analysis

Criteria	Affected Contracts	Affected CLINs	Total Contracts	Total CLINs
Initial data extraction from CADE's website	451	863	451	863
Experienced an Over Target Baseline	67	145	384	718
Did not report to at least 92.5% complete	163	348	221	370
EAC failed to reach within 10% of the final BAC	24	52	197	318
EAC failed to reach within 5% of the final BAC	29	45	168	273
Reported completion point exceeding 100%	10	19	158	254

Part I Analysis

After the CLINs that met the requirements for Part I of the analysis were organized in Table 8, the 96 CLINs were individually graphed and then all averaged into one final graph. The graphs were created in R. The x-axis is the percent complete (shown in 5% increments) and the y-axis shows the %DCAC. The black line represents the average $\%DCAC_{CPI}$ at each percent complete bucket, and the red line represents the average $\%DCAC_{SCI}$ at each percent complete bucket. After the 96 CLINs were graphed, a final graph with all averages of the 96 CLINs was graphed. If a CLIN did not have an average $\%DCAC_{CPI}$ or $\%DCAC_{SCI}$ at a certain percent complete bucket, it did not contribute to the average for that specific percent complete. Every CLIN contributed to the final average graph but not every CLIN was reported at every 5% increment. The next few figures were part of the 96 CLINs for the final averaged graph.

Figure 4 shows one of the 96 CLINs that was graphed for this research. Figure 4 shows the Army's "W58RGZ-12-C-0057, LRIP III" CLIN. For this CLIN, the status of LRIP III was reported 20 times and therefore there are 20 points along the lines for both $\%DCAC_{CPI}$ and $\%DCAC_{SCI}$. Initially, the $\%DCAC_{SCI}$, shown in red, is 5% from the

CAC, then it rises to 25% deviation at the 30% complete point, and it then regresses to 0% at the 100% complete point. The $\%DCAC_{CPI}$ is shown in black. Initially, it starts at -5% from the CAC, then it drops to -20% deviation at the 25% complete point, and then it stays steady to within 5% of CAC for the remainder of its life. Most of the 96 CLINs have a similar pattern to this CLIN: the $\%DCAC_{SCI}$ is above the $\%DCAC_{CPI}$ from beginning to end. The next two figures show atypical graphs that were part of the 96 CLINs and used for the final averaged graph.

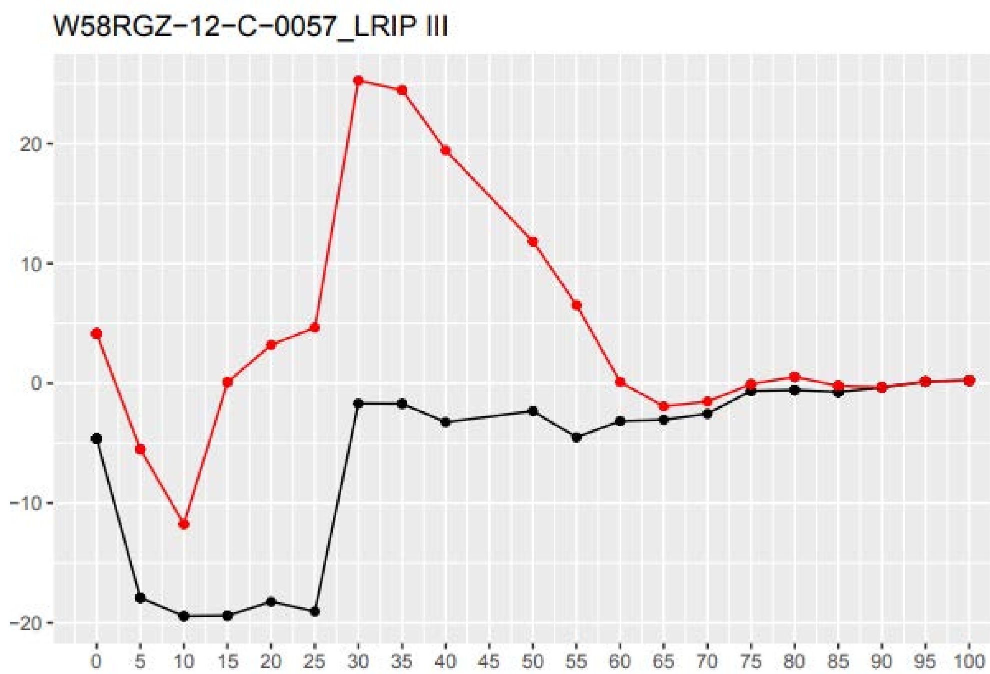


Figure 4. Contract W58RGZ-12-C-0057, LRIP III

Another CLIN is shown in Figure 5; the Navy's "N00019-03-C-0057, Pilot Production" is different from Figure 4 in that the $\%DCAC_{CPI}$ starts off above the $\%DCAC_{SCI}$ and stays higher for the remainder of the period. The $\%DCAC_{CPI}$ shows around -8% deviation at the 10% complete point whereas the $\%DCAC_{SCI}$ shows -25%

deviation at the 10% complete point. Eventually, $\%DCAC_{SCI}$ matches the path of $\%DCAC_{CPI}$ at the 80% complete point.

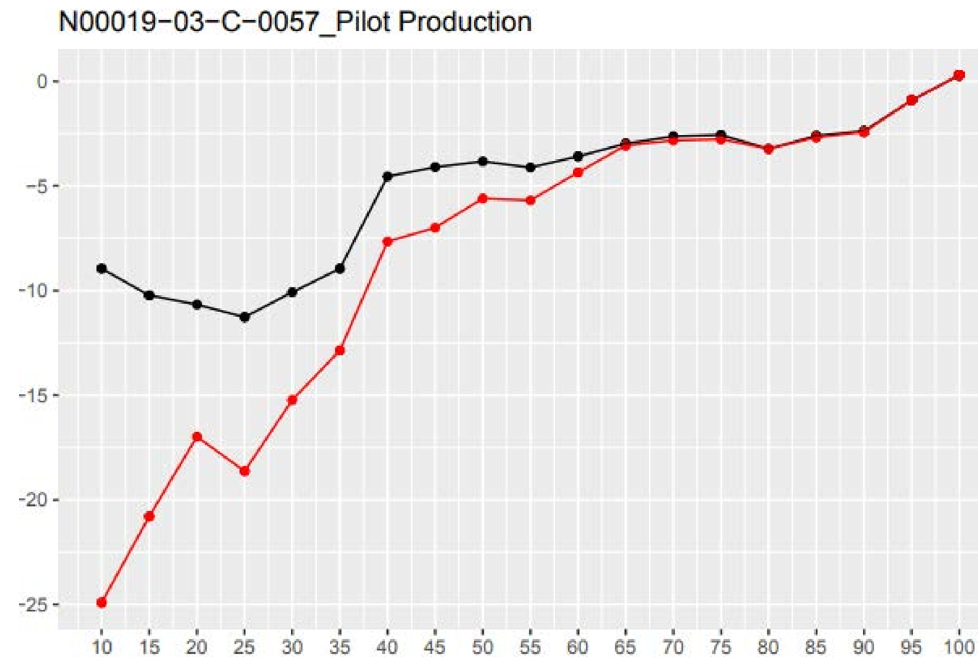


Figure 5. Contract N00019-03-C-0057, Pilot Production

Figure 6 shows the Air Force's "FA8620-10-G-3038, DO 0052" CLIN. Like Figure 4, the $\%DCAC_{SCI}$ is above the $\%DCAC_{CPI}$ at the beginning. However, at the 50% complete point, the lines cross each other as the $\%DCAC_{CPI}$ goes from having a negative deviation to a positive deviation. The lines then cross again at 60% complete, and the $\%DCAC_{SCI}$ again goes above the $\%DCAC_{CPI}$ line.

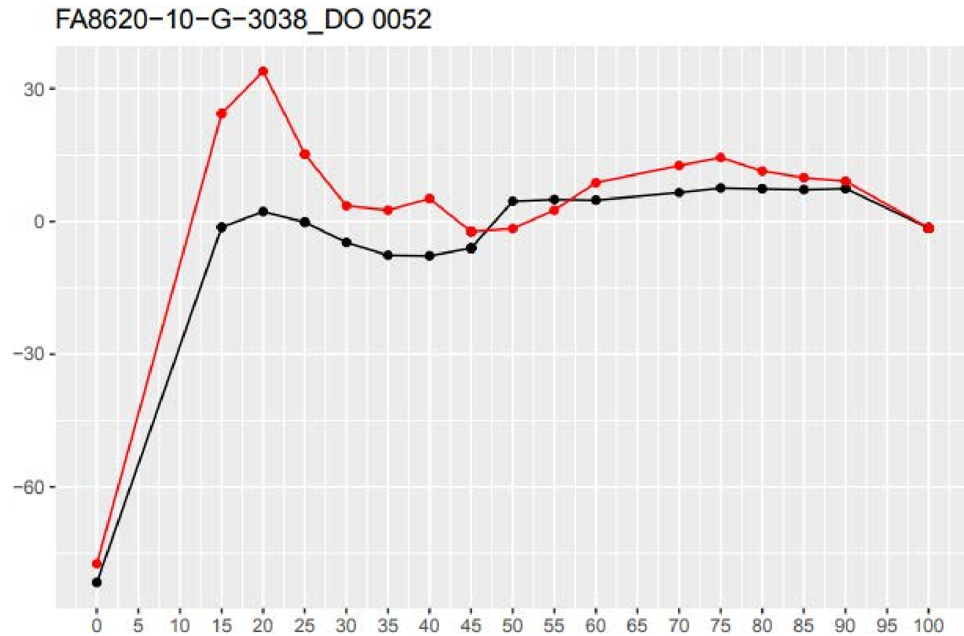


Figure 6. Contract FA8620-10-G-3038, DO 0052

After calculating and rounding the percent completes, there were noticeable differences between CLINs regarding the amount of times a contractor reported its progress. Some CLINs' numbers were sparsely reported whereas others were reported (more than) monthly. Moreover, some CLINs had longer periods of reporting because of the length/duration of the program and some were shorter. The anomalies in the amount of times a contract reported its progress was a limitation of this study because some percent completes were limited in number. Table 10 shows the number of CLINs reported between the 0% and 100% complete buckets. Some CLINs reported multiple times for the same percent complete. The first column shows the percent complete, and the second column shows the number of CLINs that were within the specified percent complete. For example, at the 5% complete point, there were 68 CLINs that reported in that percent bucket. The 68 values for $\%DCAC_{CPI}$ were averaged for the final

$\%DCAC_{CPI}$ at 5%. Likewise, the 68 values for $\%DCAC_{SCI}$ were averaged for the final $\%DCAC_{SCI}$ at 5%. The final graph with the 96 CLINS averaged is shown in Figure 7.

Table 10. Number of CLINs at Each Percent Complete

Percent Complete	Number of CLINs
0	47
5	68
10	71
15	78
20	81
25	78
30	77
35	76
40	78
45	72
50	78
55	79
60	78
65	78
70	76
75	87
80	80
85	87
90	90
95	90
100	96

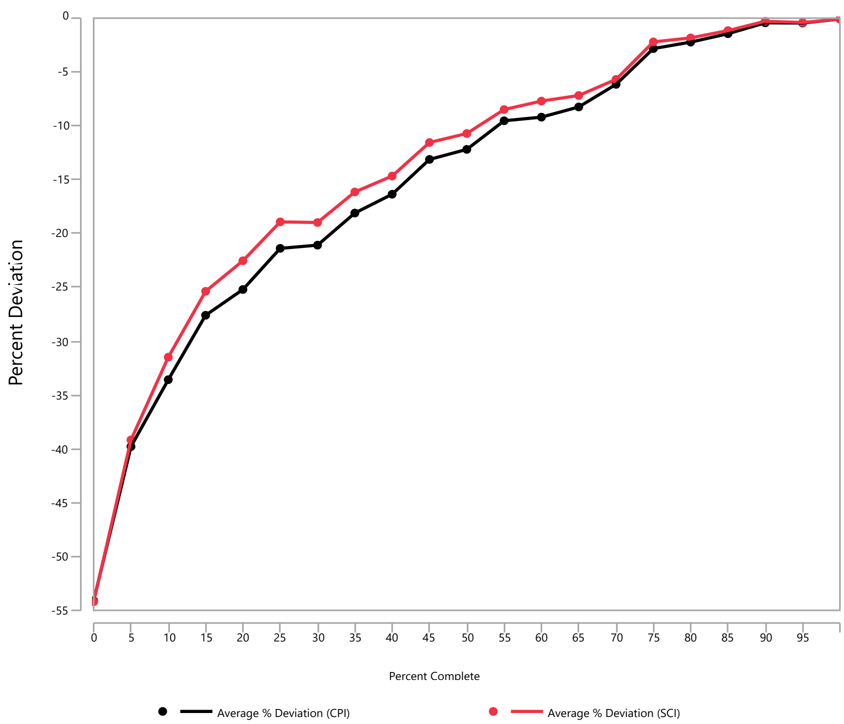


Figure 7. Average of 96 CLINs

Figure 1, in the Introduction, showed Christensen's analysis of 64 contracts from 1977 through 1996. As shown in Figure 1, EAC_{SCI} approximates the final CAC of a CLIN much quicker than EAC_{CPI}. Figure 7 shows how 21st century contracts perform with respect to today's EAC_{SCI} and EAC_{CPI} and the final cost of a contract.

When comparing Figures 1 and 7, there are two conclusions that become apparent. One, the EAC_{CPI}'s are relatively comparable with respect to when they achieve the within 10% threshold. In Figure 1, this threshold is met at approximately 50 percent complete, while modern contracts suggests this percentage is closer to 55 percent complete. The second conclusion drawn from comparing Figures 1 and 7 is that the pattern of deviations regarding EAC_{SCI} in 1996 no longer appears to hold for modern contracts. In 1996, the percent complete for when the EAC_{SCI} was within 10% of the

final cost of a contract was essentially at contract initiation. And even when narrowing this band to within 5% accuracy, a contract only needed to achieve 20% completion for the EAC_{SCI} to be extremely accurate. For modern contracts, it appears the EAC_{SCI} only achieves this within 10% accuracy band at approximately 50% completion, extending to approximately 70% completion when narrowing to 5% accuracy.

The overall conclusion is that today's EAC_{SCI} estimates closely mirror EAC_{CPI} estimates. Since SCI is the product of CPI and SPI, this suggests modern contracts maintain relatively high SPI numbers. As for empirical evidence, of the 96 CLINs in the first database, the mean SPI was 0.88, with a median value of 0.89.

Part II Analysis

The second part of the analysis calculated 95% confidence intervals to determine at what percent complete CLINs appeared to be within the +/- 5% and +/- 10% threshold of the final CAC with respect to EAC_{CPI} and EAC_{SCI}. The dataset used for the second part of the analysis looked at all CLINs that reported ≥ 92.5% complete and is shown in Table 9. The 92.5% completion point signifies completion based on research by Tracy and White (2011).

Dataset Separation

The 254 CLINs shown in Table 9 were split into two groups: one group fell in the 92.5% category and the other was the 100% category. The 92.5% category included CLINs that were between the range of ≥ 92.5% and < 99.5%. The 100% group consisted of CLINs that were ≥ 99.5% complete. There were 147 CLINs in the 92.5% category and 107 CLINs in the 100% category. None of the contracts in the 100% category were

included in the 92.5% category. The dataset in Parts II and III analyzed the absolute values of $\%DCAC_{CPI}$ ($|\%DCAC_{CPI}|$) and $\%DCAC_{SCI}$ ($|\%DCAC_{SCI}|$). The purpose of analyzing the absolute values of $\%DCAC_{CPI}$ and $\%DCAC_{SCI}$ is because it negates the occurrences of positive and negative deviations playing a role in contract stabilization. No averages were included in this portion of the analysis; instead, the actual percent completes (not bucketed percent completes) and absolute values of the actual $\%DCAC_{CPI}$ and $\%DCAC_{SCI}$ were analyzed.

This part of the research located the percent complete where a CLIN's $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$ reached 10% and stayed ≤ 10% for the remainder of the CLIN's life. For example, if a CLIN's $|\%DCAC_{CPI}|$ was ≤ 10% at the 30% point, then its $|\%DCAC_{CPI}|$ rose to >10% at the 50% point, then its $|\%DCAC_{CPI}|$ was ≤ 10% at the 70% point and the deviation from CAC remained ≤10% for the duration of its life, then the 70% complete point would be recorded as when the $|\%DCAC_{CPI}|$ reached a 10% deviation from CAC for that specific CLIN. This process was done for both the 92.5% group and 100% group for both $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$. The next part was to locate the percent complete in which a CLIN's $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$ was 5% and stayed ≤ 5% for the remainder of the CLIN's life. Sometimes the percent complete where $|\%DCAC_{CPI}|$ was at 10% deviation would equal the percent complete where $|\%DCAC_{CPI}|$ was at 5% deviation because of the lack of data points in between reporting periods. This also occurred with $|\%DCAC_{SCI}|$. Each CLIN's percent complete was recorded when it crossed the 10% and then 5% threshold for $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$. The four percent completes for when each CLIN's deviation reached 5%

and 10% for $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$ will be referred to as $\%EAC_{CPI}(5)$, $\%EAC_{CPI}(10)$, $\%EAC_{SCI}(5)$, and $\%EAC_{SCI}(10)$.

Confidence Intervals

After the percentages were found for the 5% and 10% deviations for each of the 147 CLINs in the 92.5% group and 107 CLINs in the 100% group, 95% confidence intervals were calculated for $\%EAC_{CPI}(5)$, $\%EAC_{CPI}(10)$, $\%EAC_{SCI}(5)$, and $\%EAC_{SCI}(10)$. A confidence interval provides an estimate for the mean range and creates an interval for each of the CLINs. For each of the percentages, $\%EAC_{CPI}(5)$, $\%EAC_{CPI}(10)$, $\%EAC_{SCI}(5)$, and $\%EAC_{SCI}(10)$, the two groups were 92.5% and 100%. Therefore, eight confidence intervals were created. Table 11 shows the results of the confidence intervals for the 147 CLINs in the 92.5% group. Table 12 shows the results of the confidence intervals for the 107 CLINs in the 100% group.

Table 11. Confidence Intervals for 92.5% Group

Percent Complete	$\%EAC_{CPI}(5)$	$\%EAC_{CPI}(10)$	$\%EAC_{SCI}(5)$	$\%EAC_{SCI}(10)$
Mean	74.148%	49.268%	77.273%	52.181%
Upper Confidence Limit	78.220%	54.140%	80.434%	56.632%
Lower Confidence Limit	70.076%	44.397%	74.112%	47.730%

Table 12. Confidence Intervals for 100% Group

Percent Complete	$\%EAC_{CPI}(5)$	$\%EAC_{CPI}(10)$	$\%EAC_{SCI}(5)$	$\%EAC_{SCI}(10)$
Mean	72.314%	57.085%	75.592%	62.447%
Upper Confidence Limit	77.770%	63.186%	79.971%	67.780%
Lower Confidence Limit	66.858%	50.984%	71.214%	57.114%

According to Tables 11 and 12, the confidence intervals reinforce the findings from Part I's analysis. A contract needs to be further long for EAC_{SCI} to be accurate within 5 or 10 percent of the final contract cost in comparison to those results

demonstrated by Christensen (1996). Additionally, EAC_{CPI}s appear to be reasonably comparable to those reflected in the Christensen's research with respect to what completion percentage attained when achieving either the 5% or 10% deviation from CAC. In general, both the CPI and SCI estimates appear to meet the 10% accuracy range around the 50% completion point. For the 5% accuracy range, this completion shifts to around the 75% completion point for both estimates.

Comparison Analysis

With the data from the 254 CLINs, parametric and non-parametric tests were run on the 92.5% group and 100% group. The sample t-test compares the means between the 92.5% group and 100% group to determine if there is a significant difference between the two population means. The null hypothesis is that the means of the two groups are equal and the alternative is that the two group are different. A level of significance (α) of 0.05 was used.

$$H_o: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

After the t-tests, Wilcoxon/Kruskal-Wallis tests were performed. The Wilcoxon/Kruskal-Wallis test is a nonparametric rank-based test based on comparing medians and does not assume a normal distribution. The Wilcoxon/Kruskal-Wallis test is a rank sum test which means it combines all observations and ranks them. Then, the test calculates the rank averages within each variable which calculates the test statistic. The null hypothesis is that the medians between the 92.5% group and 100% group are equal and the alternative is that the two groups are different. For the Wilcoxon/Kruskal-Wallis

test, an alpha of 0.05 was also used. The results of the t-tests and Wilcoxon/Kruskal-Wallis tests are shown in Table 13. The asterisks denote statistically significant findings at the 0.05 level of significance.

Table 13. Hypothesis Tests

	$\%EAC_{CPI}(5)$	$\%EAC_{CPI}(10)$	$\%EAC_{SCI}(5)$	$\%EAC_{SCI}(10)$
T-test	0.5865	0.0463*	0.5277	0.0036*
Wilcoxon/Kruskal-Wallis test	0.9064	0.0256*	0.7326	0.0011*

The results show that there is a significant difference between the 92.5% group and 100% group when a CLIN reaches 10% deviation from CAC. These findings are consistent with the findings from Tables 11 and 12; $\%EAC_{CPI}(10)$ in the 92.5% group met the 10% DCAC range at around the 44% complete point while $\%EAC_{CPI}(10)$ in the 100% group met the 10% range at 51%. Moreover, $\%EAC_{SCI}(10)$ in the 92.5% group met the 10% DCAC range at 48% while the $\%EAC_{SCI}(10)$ in the 100% group met the 10% range at 57%. The numbers from the 100% group met the 10% DCAC range later than the 92.5% group. This suggests that a statistically significant amount of ACWP is being documented on the later part of the EVM report for the 100% group which is not being reported for CLINs in the 92.5% group. Therefore, caution should be used with when using EACs at earlier completion percentages as reported by the 92.5% group.

Part III Analysis

The third part of the analysis created two regression models for response variables $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$ using JMP®'s stepwise function. This portion of the

research sought to differentiate large and moderate drivers of $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$ using a mixed stepwise Ordinary Least Squares (OLS) methodology to develop the models. A level of significance was set to 0.001 to determine initial predictive ability of an explanatory variable.

A population analysis uses OLS methodology to determine predictor variables that appear to be strongly associated with EAC reliability and not as a predictive model to use. In that context, data was not broken into the usual parts of a training set and a test set, and the entirety of the data was considered a characterization of the CLIN population in the aggregate. The variables used in the multiple regression model to predict $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$ come strictly from the CADE database. The CADE database provided possible predictor variables for $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$ to include the program of the CLIN, reporting contractor, contract type, program phase, dates (effective date, report from date, start date, definitization date, completion date, Estimated Completion Date (ECD), BAC date), TAB, Variance at Complete (VAC), MR, BAC, WBS Level, Military Handbook, Weapon System, and Branch. The definitions for the dates are shown in Table 14.

Table 14. Definitions of Dates

Date	Meaning
Effective Date	The end of an accounting period for a particular report.
Report From Date	The beginning of the accounting period for a particular report.
Start Date	The negotiated starting date of a contract/ when a contract is supposed to start.
Definitization Date	When the contract was definitized/when the contract was let/given to the contractor.
Completion Date	The negotiated contract completion date/when the contract is supposed to end.
ECD	The contractor's estimated completion date/when the contract is supposed to end
BAC Date	The budget at complete date/when the contract is finished being funded and no more budget will reach the contract after this point.

The following list shows the variables that were used as possible explanatory variables for the multiple regression model across all 254 CLINs.

- *MR/TAB – Continuous Variable*

  This variable shows the ratio between MR and TAB to determine what proportion of the MR was included in the TAB.

- *Money Duration – Binary Variable*

  The number of days the money of a contract flows is between the BAC date and definitization date and is also known as money flow (MF). After the number of days between the two dates was calculated, the number of days was separated into four quartiles: MF1, MF2, MF3, and MF4.

- *Contract Length – Binary Variable*

  The contract length (CL) is the number of days between the Start Date and Completion Date. After the number of days between the two dates was calculated, the number of days was separated into four quartiles: CL1, CL2, CL3, and CL4.

- *Delay Length – Binary Variable*

  The delay (D) of a CLIN is the number of days between the start date and definitization date. After the number of days between the two dates was calculated, the number of days was separated into four quartiles: D1, D2, D3, and D4.

- *Service – Binary Variable*

  The five branches/agencies in this database are Air Force (AF), Army, Navy, DoD, and MDA (Missile Defense Agency).

- *Percent Complete – Binary Variable*

  The percent complete (PC) was calculated using $BCWP_{CUM}$/BAC. All PCs that were greater than 100% were eliminated. Then the percent completes were separated into four quartiles: PC1, PC2, PC3, and PC4.

- *Commodity Type – Binary Variable*

  The 10 different types of weapons systems in the military handbook in the CADE dataset were Unmanned Aerial Vehicle (UAV), Surface Vehicle, Missile, Electronic/Automated Software, Aircraft, Ship, System of Systems, Space, Ordnance, and Other.

- *Contract Type – Binary Variable*

  The 7 types of contracts used in this analysis were Fixed Price Incentive Fee Target (FPIF), Cost Plus Fixed Fee (CPFF), Cost Plus Incentive Fee (CPIF), Cost Plus Award Fee (CPAF), Firm Fixed Price (FFP), Multiple, and Other. The Multiple variable meant that there were some CLINs that used two or more different contract types.

- *Max Reporting Level – Binary Variable*

  The WBS were from the 1.0 level to the 12.0 level for a total of twelve different levels. WBS Levels 1 and 2 were grouped together (WBS 1,2), WBS Levels 3 and 4 were grouped together (WBS 3,4), WBS Levels 5 and 6 were grouped together (WBS 5,6), WBS Levels 7 and 8 were grouped together (WBS 7,8), WBS Levels 9 and 10 were grouped together (WBS 9,10), and WBS Levels 11 and 12 were grouped together (WBS 11,12).

- *Program Phase – Binary Variable*

  The 13 different types of program phases in the dataset were Production, Service, Research Development Test & Evaluation (RDTE), Low Rate Initial Production (LRIP), Technology Development, Development, Engineering & Manufacturing Development (EMD), Sustainment, Technology Maturation & Risk Reduction (TMRR), Engineering, and Design.

Notably missing is an explanatory variable to identify contractor. Contractor was not considered because if a contractor was identified through regression analysis to be

predictive of CPI or SCI deviation percent, this association is likely tied to a particular type or characteristic of CLIN rather than an inherent systematic pattern of an individual contractor.

The following steps outline the procedures taken to determine the major and moderate drivers of $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$. Because the major and moderate drivers of $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$ were identical, the following outputs are representative for both explanatory variables.

Variance Inflation Factors (VIF)

The first part of the regression portion of this research sought to identify any predictor variables that had multicollinearity. VIF scores that are above 5 suggest that there is a linear dependency between two or more independent variables and therefore should be removed from the model. None of the independent variables had a VIF scores above 5 (shown in Figure 8).

Parameter Estimates						
Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta	VIF
Intercept	19.947856	0.691626	28.84	<.0001*	0	.
WBS1,2	9.2236924	0.772307	11.94	<.0001*	0.088024	1.1133572
WBS11,12	-6.071367	1.430855	-4.24	<.0001*	-0.03015	1.0345847
MFQ3	-1.813816	0.41035	-4.42	<.0001*	-0.03714	1.447059
MFQ4	2.3628984	0.415954	5.68	<.0001*	0.048426	1.4894424
CL1	2.4496224	0.41167	5.95	<.0001*	0.050191	1.4581965
NAVY	-2.953786	0.413462	-7.14	<.0001*	-0.06934	1.9306353
AF	-3.272188	0.423415	-7.73	<.0001*	-0.06787	1.5809501
PC1	18.9191	0.488324	38.74	<.0001*	0.366975	1.8388622
PC3	-7.613125	0.499188	-15.25	<.0001*	-0.14084	1.7479543
PC4	-14.85238	0.43295	-34.31	<.0001*	-0.34766	2.1050296
UAV	3.8055703	0.507238	7.50	<.0001*	0.057622	1.2089897
ELECTRONIC/AUTOMATED SOFTWARE	4.9281698	0.4668	10.56	<.0001*	0.082949	1.2652417
SHIP	6.5831966	0.512751	12.84	<.0001*	0.115852	1.6688092
PRODUCTION	-3.845104	0.514717	-7.47	<.0001*	-0.09054	3.0105409
SERVICE	-10.96827	2.89898	-3.78	0.0002*	-0.02746	1.079671
RDTE	-3.093039	0.512828	-6.03	<.0001*	-0.07101	2.8413201
DEVELOPMENT	9.4006373	1.401843	6.71	<.0001*	0.050221	1.1495332
EMD	4.9140576	1.341636	3.66	0.0003*	0.028584	1.2481943
SUSTAINMENT	-15.1922	3.120282	-4.87	<.0001*	-0.03473	1.0428259
ENGINEERING	-20.06425	2.566012	-7.82	<.0001*	-0.05725	1.0987371
DESIGN	-18.04033	2.88464	-6.25	<.0001*	-0.04517	1.0690161
FPIF	-5.707654	0.423801	-13.47	<.0001*	-0.11321	1.4482059
CPFF	6.035937	0.410165	14.72	<.0001*	0.116781	1.2907157
CPAF	3.1986596	0.570592	5.61	<.0001*	0.042803	1.1948956

Figure 8. Model VIF Scores and Standard Beta Coefficients

Standard Beta Coefficients

The Standard Beta coefficients associated with each of the independent variables are shown in the "Std Beta" column in Figure 8. The Standard Betas compare the strength of the independent variable to the explanatory variable. The greater the number, the stronger the effect it has on $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$. A list of the major, moderate, and minor drivers is shown in Table 15 from the variables shown in Figure 8.

Table 15. Significant Regression Variables

Variable	Definition	Overall effect on CAC Estimate Reliability
PC1	Contracts in the first quartile of percent complete (0% to 24.99%)	Strongly negative
PC4	Contracts in the fourth quartile of percent complete (75% to 100%)	Strongly positive
PC3	Contracts in the third quartile of percent complete (50% to 74.99%)	Strongly positive
CPFF	Cost Plus Fixed Fee (CPFF) contract	Moderately negative
Ship	Navy surface ship contract	Moderately negative
WBS 1, 2	Contracts with WBS level 1 or 2	Moderately negative
Electronic/Automated Software	Electronic/Software contracts	Moderately negative
FPIF	Fixed Price Incentive Fee Target (FPIF) contract	Moderately positive
UAV	An unmanned aerial vehicle (UAV) contract	Minorly negative
Development	A contract solely involving development	Minorly negative
CL1	Contracts that have less than 963 days from the contract start date to completion date	Minorly negative
MFQ4	Contracts that have between 1705 and 7765 days from the definitized date to the BAC date	Minorly negative
CPAF	Cost Plus Award Fee (CPAF) contract	Minorly negative
EMD	Engineering & Manufacturing Development contract	Minorly negative
Engineering	A contract solely involving engineering	Minorly positive
AF	A contract in the Air Force	Minorly positive
Production	A contract solely involving production	Minorly positive
Navy	A contract in the Navy	Minorly positive
Design	A contract solely involving design	Minorly positive
RDTE	A contract in the Research Development Test & Evaluation phase	Minorly positive
Sustainment	A contract solely involving sustainment	Minorly positive
MFQ3	Contracts that have between 913 and 1704 days from the definitized date to the BAC date	Minorly positive
WBS 11, 12	Contracts with a WBS extending to level 11 or 12	Minorly positive
Service	A contract solely involving service	Minorly positive

Cook's Distance Test

The Cook's Distance detects overly influential data points that could possibly skew the results. Typically, if a Cook's D value is greater than 0.5, the data point(s) are justified in removal from the dataset. Because of the large sample size (over 10,000 rows of data), some outliers are expected. However, this number was changed to 0.004 in keeping with the general spirit of 4/*n*, where *n* is the number of data points (Bollen & Jackman, 1990). The new bar was 0.0004 but because this would result in flagging too many data points, the number was capped at 0.004. With this new bar of 0.004, 27 data points were excluded from the analysis. With the removal of the 27 points, the Cook's D plot is shown in Figure 9.

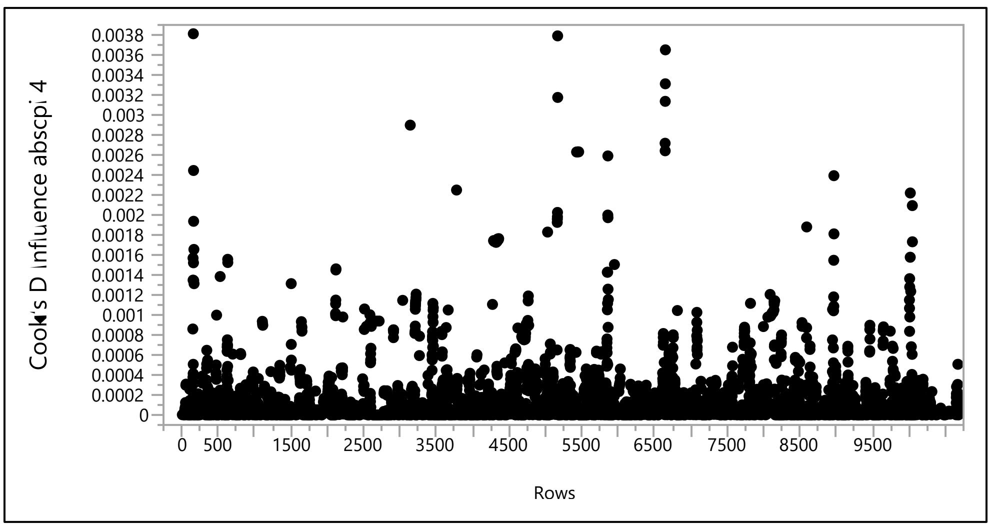


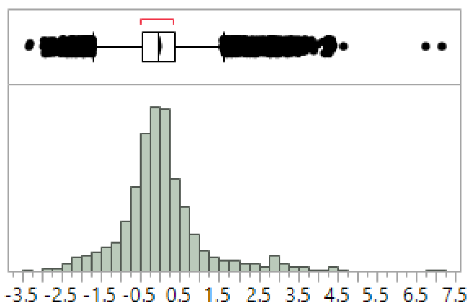
Figure 9. Display of Cook's D Plot

Studentized Residuals

Like the Cook's D test, a histogram of the studentized residuals identifies potential outliers in the data set. The histogram is shown in Figure 10. The status quo of analyzing studentized residuals is to see if they are within 3 standard deviations above or

below the standard normal distribution’s mean of zero; this keeps the assumption of a normal distribution of the residuals. Because of the large dataset, it was assumed that there would be residuals that would go beyond the 3 standard deviations. The points that went past the 3 standard deviations were not removed from the dataset.

Studentized Resid abscpi



Quantiles

100.0%	maximum	7.1837218191
99.5%		3.7612995584
97.5%		2.8269858264
90.0%		1.0087742047
75.0%	quartile	0.3571022169
50.0%	median	-0.066451699
25.0%	quartile	-0.474308143
10.0%		-1.067198199
2.5%		-1.945680293
0.5%		-2.388356773
0.0%	minimum	-3.353195239

Summary Statistics

Mean	-7.73e-6
Std Dev	1.0001254
Std Err Mean	0.0096876
Upper 95% Mean	0.0189818
Lower 95% Mean	-0.018997
N	10658

Figure 10. Studentized Residuals

Shapiro-Wilk (S-W) and Breusch-Pagan (B-P) Tests

The models for $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$ must pass the assumptions of being normally distributed and possessing constant variance for a regression analysis. However, because these models are being used to identify major and moderate drivers of the explanatory variables (as shown in Table 15), the following tests were not of major concern for the population analysis. The R^2 value for the model was 0.4857.

The Shapiro-Wilk (S-W) goodness of fit test determines whether a random sample comes from a normal distribution and is shown in Figure 11. The null hypothesis is that the model residuals possess a normal distribution and the alternative is that they do not. When the S-W test was performed on the model for $|\%DCAC_{CPI}|$, the test for normality failed statistically. However, the graph still shows a normal distribution with

most of the residuals in the middle of the curve. For this reason, this is considered a "soft fail" and the test for normality passes.

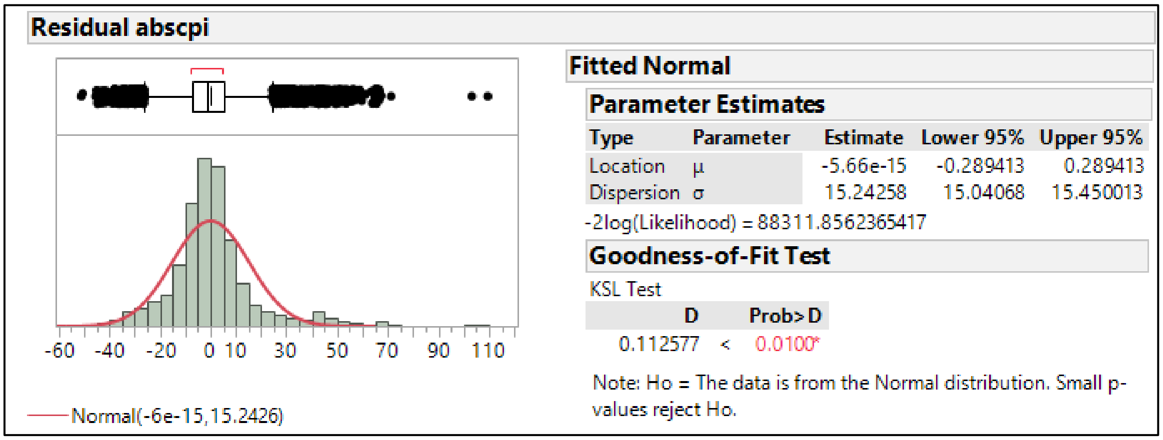


Figure 11. Shapiro-Wilk Test Results

Next, the Breusch-Pagan (B-P) tests the assumption of constant variance of the error term. This test is used with the purpose of identifying whether heteroscedasticity is present in the model. In order to pass the assumption of constant variance, the p-value from the test must be above 0.05. The null hypothesis is that the model's assumption of constant variance holds and the alternative is that it does not. Table 16 shows that constant variance is not shown within the residuals.

Table 16. Breusch-Pagan Test Results

<table>
<tr><th></th><th></th><th>B-P Test Statistic</th><th>P-Value</th></tr>
<tr><td>Sample Size</td><td>10,658</td><td rowspan="4">9017.790</td><td rowspan="4">0.0000000</td></tr>
<tr><td>Model Degrees of Freedom</td><td>24</td></tr>
<tr><td>SSE</td><td>2,476,007.3</td></tr>
<tr><td>SSR</td><td>973,380,290</td></tr>
</table>

When analyzing the residuals versus predicted plot in JMP®, the figure shows that the test for constant variance fails (see Figure 12). Since this analysis does not use the

regression models as a predictive tool but to identify key drivers for the response variable, it does not consider this fail to be a major impact of the analysis.

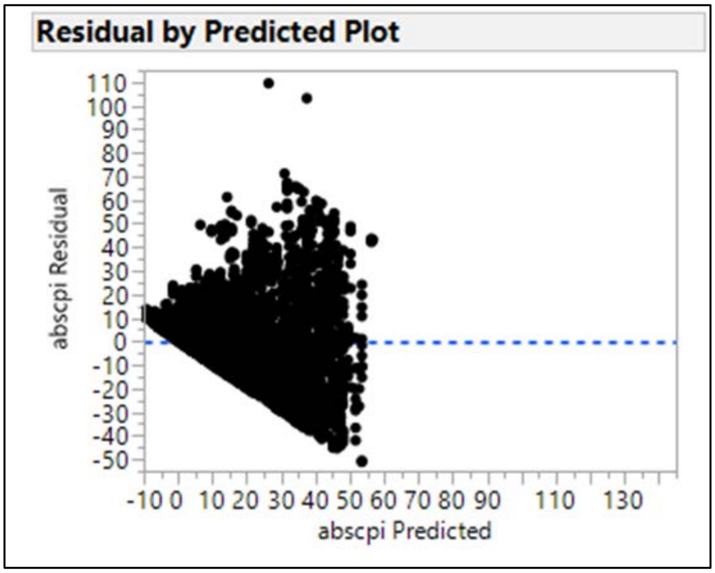


Figure 12. Residuals by Predicted Plot

Conclusion

This chapter discussed the methodology as well as the analysis and results at each step of the process. There were three parts of the analysis: the first part replicated Christensen's work from 1996, the second part re-analyzed a portion of Tracy and White's work, and the third part identified major and moderate drivers for $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$. The first part of the analysis shows that Christensen's study on contract stability is no longer applicable to modern day contracts. When comparing Figures 1 and 7, it is apparent that the pattern of deviations regarding EAC_{SCI} in 1996 no longer appears to hold for modern contracts. For modern contracts, it appears that EAC_{SCI} achieves 5% accuracy at the 70% complete point whereas Christensen's 5% accuracy for EAC_{SCI} was achieved at the 20% complete point.

For the second and third parts of the analysis, the absolute values of $\%DCAC_{CPI}$ and $\%DCAC_{SCI}$ were taken to negate the occurrence of positive and negative deviations playing a role in contract stabilization. The second part of the analysis concludes that CLINs in the 92.5% cohort and 100% cohort are statistically different at the 10% threshold. This suggests that a statistically significant amount of ACWP is being documented on the last EVM report for the 100% cohort. Lastly, the third part of the analysis identified major and moderate drivers of $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$. The main drivers of |%DCAC| are the percent complete. Contracts in the first quartile of percent complete have a strongly negative effect on CAC estimate reliability while contracts in the third and fourth quartiles of percent complete have a strongly positive effect on CAC estimate reliability. This makes sense because estimates become substantially more accurate as contracts get closer to the end stages. The variables that have a moderately negative effect on CAC, according to Table 15, are CPFF contracts, contracts dealing with Navy surface ships (Ship), contracts with a WBS level of 1 or 2 (WBS 1,2), and Electronic/Automated Software contracts. FPIF had a moderately positive effect on CAC.

IV. Journal Article

Overview

This chapter serves as the journal article that was submitted to the *Journal of Public Procurement* by the researchers, Dr. Edward White, Dr. Dan Ritschel, Mr. Chad Millette, and 1st Lt Deborah Kim. The journal article is a condensed version of the thesis.

Reliability of Estimates at Completion for Department of Defense Contracts

Introduction

Earned Value Management (EVM) is an amalgamation of business practices that provides a structured method to measure and to analyze performance. Proper interpretation and application of EVM measures serve as a monitoring tool for project managers on the status of their programs in the categories of cost, schedule, and performance. Specifically, EVM provides a way to organize project schedule, budget, and planning components that can produce forecasts and status determinations. The basis for project alterations necessary to meet established goals originates from a comparison of the current state of a program to the forecasted measure. EVM data provides inputs for these forecasts.

These forecasts, termed Estimates at Completion (EACs), are used to predict the final cost of a program. Often, EACs calculated by the government and contractors are significantly lower than what finalizes as the last contract cost. Pressure from

stakeholders to keep programs from being cancelled has led to a toxic culture of reporting optimistically low EAC calculations (Christensen, 1996). Organizations encourage goals that are unrealistic because there is an over optimism that these goals are attainable. This mentality has resulted in programs that cost more than planned and produce results that do not satisfy all requirements (GAO, 2009).

Within EVM, two particular performance indices are often used to estimate EAC in order to monitor how well a contract is pacing with respect to planned budget and schedule. These metrics are the Cost Performance Index (CPI) and the Critical Ratio (CR) (Lewis, 2001; Anbari, 2003; DAU, 2018). CR is the product of the CPI and the Schedule Performance Index (SPI). According to Christensen's analysis (1996), estimating the final Cost at Completion (CAC) of a contract using EAC_{CR} is a quicker predictor of the actual final cost versus using EAC_{CPI}. [Note: Christensen referred to the CR as the Schedule Composite Index (SCI)]. When Christensen used CR as the performance indicator to calculate EAC (EAC_{CR}), there was less than a 5% deviation from the final cost at a contract's 20% complete point. However, when Christensen used CPI as the performance indicator to calculate EAC (EAC_{CPI}), there was less than a 5% deviation from the final cost at a contract's 70% complete point.

Using CR as the performance indicator seemed a quicker method of calculating the actual final CAC according to Christensen (1996) and therefore the EAC appeared more reliable sooner in a contract's life. This analysis was based on 64 contracts available in 1996 and has become routinely cited by subsequent EVM authors and academic literature (Tracy and White, 2011; Petter et al., 2015). The salient question:

does this still hold for Department of Defense (DoD) contracts in the 21st century? This paper replicates Christensen's analysis to not only answer this question but to also investigate a contract's progression to ascertain when the EAC is a reliable estimate of a contract's final cost by using either CPI or CR.

Database and Methodology

To combat and possibly militate against cost growth, Congress enacted the Weapon System Acquisition Reform Act of 2009, often called WSARA as public law 111-23. The act created a Pentagon office - Office of Cost Assessment and Program Evaluation (CAPE). CAPE's mission is to provide independent program analyses and insights as requested by the Under Secretary of Defense for Acquisition, Technology and Logistics and Congress. Additionally, CAPE reviews programs that may be, or already are, struggling in the acquisition process. To facilitate their mission, CAPE initiated the development of the Cost Assessment Data Enterprise (CADE), the Department's initiative to identify and integrate data from disparate databases and systems for better decision-making, management of, and oversight of the Department's acquisition portfolio.

CADE is the central repository for all Acquisition Category I (ACAT I) EVM data and houses, among many things, the Integrated Program Management Reports (IPMRs). The IPMR is a contractually required report, prepared by the contractor, containing performance information derived from the internal EVM System. IPMRs are submitted by the program offices and have monthly and quarterly updates for all the

Work Breakdown Structures (WBSs) for a project to include data on the cost and schedule of programs on major defense contracts.

EVM is required for all cost or incentive contracts equal to or greater than $20 million and/or have a high risk in development work for the government (Department of the Air Force, 2007). All DoD cost or incentive contracts that exceed $20 million require compliance with ANSI/EIA-748 per DoD 5000.2 (EVM, 2017). Because of this dollar threshold, no contracts under $20 million were used for any analysis in this paper. All data were extracted from CADE's integrated web-based site that works as a centralized virtual library. This research used only completed or nearly-completed contracts to predict EAC and therefore no programs initiated in 2017 were included in this study.

A particular contract in the study's database may have more than one Contract Line Item Number (CLIN). CLINs are specified in the Federal Acquisition Regulation (FAR) part 4.10 and serve two purposes. First, they break the contract down by the commodities being procured (labor hours of services, funding for travel, quantity of products, etc.). Secondly, they provide for traceable accounting classification citations. For the study's database, each separate CLIN was considered a unique entity and EVM data was obtained at that level.

From the initial data pull from CADE, the data is filtered through a series of exclusion criteria. These include contracts or CLINs considered Over the Target Baseline (OTB), too far along when initially tracked with respect to percent complete (beyond 20%), not reporting at least 92.5% percent complete (in keeping with Tracy and White (2011)), and those CLINs not converging within 5% of the Budget at Complete

(BAC) for the last reported EVM data. Essentially, the final database consists of contracts, and consequently CLINs, that represent a fairly complete snapshot from initiation to completion at the 1.0 WBS level of EVM data availability from 2000-2017 so that the analysis can accurately capture statistical trends over time.

Three separate stages comprise the analysis and subsequent results in this paper. The first involves duplicating the work of Christensen (1996) to ascertain if trends in 21st century DoD contracts mirrors those of the past. To facilitate the preparation of the graphs, each deviation (as calculated from either EAC_{CPI} or EAC_{CR} from the last reported BAC) was assigned to discrete stages of percent complete starting at 0% and ending at 100%, increasing in increments of 5%. Similar to that of Tracy and White (2011), a +/- 2.5% margin was used to create these incremental buckets. Any percentage within that margin was bucketed into the closest 5%. For example, the 20% bucket included any deviation equaling 17.6% to 22.4%. Equations (1) through (5) reflect the requisite calculations involved. Cumulative EVM metrics are used for a CLIN at a given percent complete. Table 17 lists and defines the relevant EVM metrics used (see DAU (2017) for more information).

$$\text{Percent complete} = BCWP_{CUM}/\text{Final BAC} \quad (1)$$

$$EAC_{CPI} = ACWP_{CUM} + [(\text{Final BAC} - BCWP_{CUM})/CPI_{CUM}] \quad (2)$$

$$EAC_{CR} = ACWP_{CUM} + [(\text{Final BAC} - BCWP_{CUM})/(CPI_{CUM} * SPI_{CUM})] \quad (3)$$

$$CPI_{Dev\%} = (EAC_{CPI} - \text{Final ACWP})/\text{Final ACWP} \quad (4)$$

$$CR_{Dev\%} = (EAC_{CR} - \text{Final ACWP})/\text{Final ACWP} \quad (5)$$

[Insert Table 17]

The second stage of analysis involves calculating 95% confidence intervals to determine at what percent complete did CLINs appear to be within the +/- 5% and +/- 10% threshold of the final ACWP (referred to as Actual Cost in Industry) with respect to EAC_{CPI} or EAC_{CR}. In addition, hypothesis tests were completed to determine equivalency of results for those CLINs completing at approximately 100% (99.5% or higher rounded to 100%) versus those completing at 92.5% or higher but less than 100%. This was done to determine if Tracy and White's (2011) conclusions still held with respect to completion equivalency. These hypothesis tests include t-tests for comparing means, which assumed unequal variances, as well as Wilcoxon rank-sum tests to test equivalency of medians. The large sample sizes (described in the next section) did not necessitate a non-parametric test, however, these were conducted as a cross-check of parametric results. For all tests, a 0.05 level of significance was chosen.

The last part of the analysis involved performing an Ordinary Least Squares (OLS) analysis to characterize major, moderate, and minor influencers on EAC reliability as denoted by either equation (4) or (5). This was done for both the absolute values of the $CPI_{Dev\%}$ and $CR_{Dev\%}$ responses, denoted $AbsCPI_{Dev\%}$ and $AbsCR_{Dev\%}$, respectively. The aim of the OLS analysis focused on identifying a graduated scale of predictors that appeared to be strongly associated with EAC reliability and not as a predictive model to use. In that context, data was not broken into the usual parts of a training set and a test

set, and the entirety of the data was considered a characterization of the CLIN population in the aggregate.

With respect to identifying possible explanatory variables for consideration, Poulos and White (2010), Tracy and White (2011), and Trudelle et al. (2017), to include cited references within, document several potential variables as possible predictive factors. Table 18 highlights the type of variables considered in the OLS analysis. Notably missing is an explanatory variable to identify contractor. Contractor was not considered because if a particular contractor was identified through regression analysis to be predictive of CPI or CR deviation percent, this association is likely tied to a particular type or characteristic of CLIN rather than an inherent systematic pattern of an individual contractor.

[Insert Table 18]

Consistent with any other OLS approach, multicollinearity, outliers, and influential data points are investigated in order to prevent variable selection bias. Variance Inflation Factors (VIF) highlight the linear relationship between independent variables and a VIF score higher than 5 suggests multicollinearity. Regarding outliers, any studentized residual greater than three standard deviations is categorized as a possible source of concern. However, due to the large dataset size (over 10,000 rows of data) some outliers are expected. Lastly, Cook's Distance detects overly influential data points possibly skewing the results. Again because of the large sample size, the customary

number of 0.5 was capped at 0.004 in keeping with the general spirit of 4/*n* (Bollen and Jackman, 1990), where *n* is the number of datapoints, but not dropping it as far as 0.0004, which would unreasonably flag too many points. Lastly, the customary residual assumptions of normality and constant variance were investigated for overall robustness of the final OLS analysis.

For both OLS analyses, a mixed stepwise procedure is adopted to arrive at the results presented in the next section. A level of significance is set to 0.001 to determine initial predictive ability of an explanatory variable. From there, the preliminary selected variables are investigated to determine their practical effect on the respective model. If a particular explanatory variable is determined to have less than a 1% relative effect on a particular model's response, then that variable is excluded from the list of associative findings. This is done to minimize a variable being statistically significant but having little practical effect. All analysis in this article used JMP12 Pro, Excel or R.

Results and Analysis

As of 26 July 2017, 451 contracts consisting of 863 CLINs were obtained from the CADE's integrated web-based site. Table 19 highlights from the beginning to the end the exclusion criteria used to create the requisite database for the first part of the data analysis. The largest exclusionary criteria evident in Table 19 consists of CLINs that had not reported to 99.5% or higher completion. Of all the CLINs removed, this lack of completion accounted for approximately 77% of the removals. Since the aim of the first

part of the analysis is to comparably replicate the work of Christensen (1996), such removals were required.

[Insert Table 19]

Figure 13 shows the graph as presented by Christensen (1996). As concluded in his work, EAC_{CR} approximates the final cost of a contract much quicker than EAC_{CPI}. For comparison, Figure 14 shows how 21st century contracts perform with respect to today's EAC_{CR} and EAC_{CPI} and the final cost of a contract. As with Christensen, the points on the graph represent averages of the CLINs' deviations from the final cost in each percentile completion bucket. When comparing the figures, two conclusions become apparent. One, the EAC_{CPI}'s are relatively comparable with respect to when they achieve the within 10% threshold as selected by Christensen (1996). In 1996, the percentage was approximately 50 percent complete, while modern contracts suggests this percentage is closer to 55 percent complete.

[Insert Figures 13 and 14]

The second conclusion drawn from comparing Figures 13 and 14 is that the pattern of deviations regarding EAC_{CR} in 1996 no longer appears to hold for modern contracts. In 1996, the percent complete for when the EAC_{CR} was within 10% of the final cost of a contract was essentially at contract initiation. And even when narrowing

this band to within 5% accuracy, a contract only needed to achieve 20% completion for the EAC_{CR} to be extremely accurate. For modern contracts, it appears the EAC_{CR} only achieves this within 10% accuracy band at approximately 50% completion, increasing to approximately 70% completion when improving to 5% accuracy. The overall conclusion is that today's EAC_{CR} estimates closely mirror EAC_{CPI} estimates. Since CR is the product of CPI and SPI, this suggests modern contracts maintain relatively high SPI numbers. As for empirical evidence, of the 96 CLINs in the first database, the mean SPI was 0.88, with a median value of 0.89.

The second stage of the analysis calculates 95% confidence intervals to ascertain at what completion percentage do CLINs appear to be within the +/- 5% and +/- 10% threshold of the final CAC with respect to EAC_{CPI} or EAC_{CR}. [Note: the final CAC is taken as the last reported cumulative ACWP for a given contract.] In addition, hypothesis tests are conducted to determine if the equivalency results demonstrated by Tracy and White (2011) still hold for those CLINs completing at approximately 100% (99.5% or higher rounded to 100%) versus those completing at 92.5% or higher but less than 99.5%. [Note: for communicative ease, these two groups will be referred to as the 92.5% cohort and the 100% cohort, respectively.] For both sets of inferential analysis, a less restrictive database was developed from the starting 451 contracts and 863 CLINs.

Table 20 lists the inclusion and exclusion criteria that not only forms the basis for the second set of analysis, but it also serves as the database for the regression analysis in the third part of the results. Of the 254 CLINs, 147 completed at or greater than 92.5% but less than 99.5%, while the remaining 107 reported a completion percentage of 99.5%

or greater. Table 21 shows the 95% confidence intervals for both the 92.5% and 100% cohorts, while Table 22 shows both the parametric and non-parametric tests for comparing equivalency between these groups.

[Inserts Table 20-22]

With respect to Table 21, the confidence intervals reinforce the findings from the first part of this section. A contract needs to be further long for EAC_{CR} to be accurate within 5 or 10 percent of the final contract cost in comparison to those results demonstrated by Christensen (1996). Additionally, EAC_{CPI}s appear to be reasonably comparable to those reflected in the 1990's with respect to what completion percentage attained when achieving either the 5% or 10% accuracy band. In general, both the CPI and CR estimates appear to meet the 10% accuracy range around the 50% completion point. For the 5% accuracy range, this completion shifts to around the 75% completion point for both estimates.

Drilling down more, these overall conclusions do hold for both the 92.5% and 100% contract cohort groups for the 5% threshold, but there is some statistical variation when it comes to specific conclusions based on these groups and the 10% accuracy threshold. As seen in Table 22 from the various parametric and non-parametric tests, the 147 contracts within the 92.5% contract cohort and the 107 contracts within the 100% contract cohort are not statistically equivalent with respect to the 10% accuracy threshold at the 0.05 level of significance. Specifically, as shown in Table 21, the 92.5% cohort

suggests possibly meeting this accuracy range at around the 44% completion point using EAC_{CPI} and around the 48% completion point using EAC_{CR}. For the 100% cohort, these percentages increase to 51% and 57%, respectively. This suggests that a statistically significant amount of ACWP is being documented on the last EVM report for this cohort, which is not being reported for CLINs in the 92.5% cohort. Therefore, caution is warranted in using earlier completion percentages as reported by the 92.5% cohort group.

For the third and final analysis, the results pertain to documenting a population analysis using the OLS method to differentiate large and moderate predictors of $AbsCPI_{Dev\%}$ and $AbsCR_{Dev\%}$ from minor, but still statistically significant, predictors. Given the results from the first part of the analysis that suggest modern contracts have relatively high SPI values, shared common predictors for $AbsCPI_{Dev\%}$ and $AbsCR_{Dev\%}$ are expected. The database used for this third part mirrors that of the second part and involved 254 CLINs, as highlighted in Table 20. For the regression analysis, the starting database consisted of 10,686 rows of data.

Beginning with the $AbsCPI_{Dev\%}$ response, 18 rows were excluded due to these responses having a value larger than 100%. This exclusion was performed to militate against these points from affecting the OLS output and only constituted a 0.17% loss of data. Table 23 reflects the OLS predictors and associated p-values; all are highly significant from at least the 0.0001 level of significance. Relative effects are based upon the standardized regression coefficients. Four data points were excluded due to relatively high Cook's D. The presented variables in Table 23 explain approximately 46% of the $AbsCPI_{Dev\%}$ response variability. The largest VIF score, studentized residual, and Cook's

D value were approximately 1.94, 4.4, and 0.0032, respectively. Lastly, the residuals presented with an approximate normal distribution, centered around zero with evidence of heteroscedasticity. Since the aim is not to use as a predictive model, but more of a discriminator of explanatory variables, the non-constancy was not an issue. [Note: transforming using the natural logarithm did alleviate this issue, but more importantly the same predictive variables were still significant.]

[Insert Table 23]

Regarding the $AbsCR_{Dev\%}$ response, a similar trend of results occurred. For this response 56 rows were excluded due to these responses having a value larger than 100%. This exclusion constituted a 0.52% loss of data. Table 24 reflects the OLS predictors and associated p-values; again, all are highly significant from at least the 0.0001 level of significance. As before, relative effects are based upon the standardized regression coefficients. Four data points were excluded due to relatively high Cook's D. The presented variables in Table 24 explain approximately 47% of the $AbsCR_{Dev\%}$ response variability. The largest VIF score, studentized residual, and Cook's D value were approximately 1.81, 5.6, and 0.0031, respectively. Lastly, the residuals presented with an approximate normal distribution, centered around zero with again evidence of heteroscedasticity.

[Insert Table 24]

The final conclusions between Tables 23 and 24 (Table 25 contains the definitions of all the predictive significant variables) are very consistent with just minor deviations within the lower significance spectrum. This is not surprising given the results of stage one and two of the analysis. With respect to overall trends, the same five variables dominate with respect to determining what is statistically associated with the percent deviation from the final cost of a contract using either EAC_{CPI} or EAC_{CR}.

[Insert Table 25]

By a large margin, the percentile completion quartiles are a very strong predictor of how accurate either EAC is with determining a contract's final CAC. As a contract nears completion, this accuracy increases. In general, as a contract nears the 50% completion point, accuracy of the CAC is approximately off by 16% to 20%, while completing near the 75% point lowers this inaccuracy to between 8% to 12%. This is in keeping with previous results.

Two new findings are the moderate significant effects of Cost-Plus-Fixed-Fee (CPFF) contracts as well contracts with reporting to WBS level 2 or higher. Both of these variables negatively affect either EAC in estimating a contract's final CAC. Approximately 21% of the CLINs in the database identified as strictly being CPFF. For those cases, and on average, the deviation percentage between the EAC and the final CAC increased by 6%. For those contracts where the reporting is not below WBS level

2, the deviation also increased. On average, those experienced a decrease of accuracy by around 10% to 11%.

Discussion and Conclusion

One common theme permeates throughout the three stages of analysis presented previously. That is, the main driver of determining how well EAC_{CPI} or EAC_{CR} predicts a contract's final cost stems from how mature the contract is with respect to completion. A program manager overseeing a 21st century contract should place very little reliability on how either the EAC_{CPI} or EAC_{CR} tracks the contract's final cost at the beginning to even through the mid-way point. Regarding the EAC_{CPI}, such a conclusion is similar to what Christensen (1996) reported years ago. In contrast, this is very much different than what was observed during the 1990's with respect to EAC_{CR}.

Christensen (1996) reported that EAC_{CR} appeared to have a very high accuracy of estimability of the CAC such that at even the 20% completion point, the relative error between EAC_{CR} and CAC was less than 5%. For 21st century contracts, that percentage has greatly increased to around 70-75%, which is closely aligned to the reliability performance of EAC_{CPI} both in the past and present. Therefore, a program manager should turn to other parametric tools to predict a contract's final cost from contract initiation to around the mid-way point of a contract's completion. Tracy and White (2011) present various models to provide such estimates.

In addition to percent complete, two other factors contributed to increasing or decreasing the accuracy of estimating the final cost of a contract by either using EAC_{CPI} or EAC_{CR}. Unlike completion percentage, CPFF contracts and those where the reporting is not beyond WBS level 2 negatively affect accuracy. Both of these findings are not unexpected. Cost-type contracts/projects are more risky and assume more uncertainty than fixed-type projects by their very nature. Therefore, a contract that is deemed CPFF possesses a selection bias to exhibit a greater deviation from an EAC to the CAC throughout its progression. A similar reasoning can be deduced by those contracts that are only reported at level 2 or higher of the WBS. Being unable to report beyond a level 2 WBS suggests less details in the reported EVM data. This lack of granularity might likely increase the chance of having cost growth and/or schedule slippage, which in turn effects the CPI and SPI.

Although the focus of this paper is on United States DoD contracts, there are other areas of relevance with respect to civilian application as well as confirmation of findings. EVM is utilized by procurement officials in a multitude of countries such as Australia, Brazil, Japan and Sweden (Antvik, 2001; Marshall et al., 2008) and found in a wide array of project types from construction to software development (Bhosekar and Vyas, 2012). 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. This also pertains to further investigating whether modern contracts in other fields and areas also appear to have relatively high SPI values as suggested by the data within the CADE database used for this paper’s analysis.

References

Anbari, F.T. (2003), "Earned value project management method and extensions", *Project Management Journal*, Vol. 34 No. 4, pp. 12-23.

Antvik, S. (2001), "Why was earned value management so important to the Swedish government in the Gripen project?", *The Measurable News*, September, pp. 1, 18-19, 21, 25 and 28.

Bhosekar, S.K. and Vyas, G. (2012), "Cost controlling using earned value analysis in construction industries", *International Journal of Engineering and Innovative Technology*, Vol. 1 No. 4, pp. 324-332.

Bollen, K.A. and Jackman, R.W. (1990), "Regression diagnostics: an expository treatment of outliers and influential cases", in Fox, J. and Long, J.S. (Eds.), *Modern Methods of Data Analysis*, Sage, Newbury Park, CA, pp. 257-91.

Christensen, D.S. (1996), "Project advocacy and the estimate at completion problem", *The Journal of Cost Analysis*, Vol. 13 No. 1, pp. 35-60. doi:10.1080/08823871.1996.10462298

Defense Acquisition University (2017), “Earned Value Management ‘Gold Card’”, available at: https://www.dau.mil/tools/Lists/DAUTools/Attachments/19/20170720-gold%20card%201%20sided.pdf (accessed 3 February 2018).

Department of the Air Force (2007), *Air Force Cost Analysis Handbook*, Author, Washington, DC.

Earned Value Management (2017), “ANSI EIA 748 Earned Value Management”, available at: http://acqnotes.com/acqnote/tasks/ansi-eia-748-earned-value-management (accessed 12 February 2018).

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

Lewis, J.P. (2001), *Project planning, scheduling, and control: a hands-on guide to bringing projects in on time and on budget*, (3rd ed.), McGraw-Hill, New York, NY.

Marshall, R.A., Ruiz, P., and Bredillet, C.N. (2008), “Earned value management insights using inferential statistics”, *International Journal of Managing Projects in*

Business, Vol. 1 Issue 2, pp. 288-294, https://doi.org/10.1108/17538370810866386

Petter, J.L., Ritschel, J.D. and White, E.D. (2015), “Stability properties in Department of Defense contracts: answering the controversy”, *Journal of Public Procurement*, Vol. 15 No. 3, pp. 341-364.

Poulos, E.T. and White, E.D. (2010), “Using growth models to improve accuracy of estimates at completion for over target baseline contracts”, *Journal of Cost Analysis and Parametrics*, Vol. 3 No. 2, pp. 37-43, DOI:10.1080/1941658X.2010.10462233

Tracy, S.P. and White, E.D. (2011), “Estimating the final cost of a DoD acquisition contract”, *Journal of Public Procurement*, Vol. 11 No. 2, pp. 190-205.

Trudelle, R., White, E.D., Koschnick, C., Ritschel, J.D. and Lucas, B. (2017), “Estimating an acquisition program’s likelihood of staying within cost and schedule bounds”, *Defense Acquisition Research Journal*, Vol. 24 No. 4, pp. 600-625. Available at https://doi.org/10.22594/dau.17-775.24.04 (accessed 12 January 2018).

Table 17: Primary Earned Value Management metrics used.

EVM Metric	Definition	Formula
BCWP (EV in Industry)	Budgeted Cost of Work Performed (Earned Value)	Sum of the budgeted cost of all completed work packages
ACWP (AC in Industry)	Actual Cost of Work Performed (Actual Cost)	Sum of the actual costs of all completed work packages
BCWS (PV in Industry)	Budgeted Cost for Work Scheduled (Planned Value)	Sum of the budgeted cost of all work packages scheduled
CPI	Cost Performance Index	Cost efficiency of a program. Calculated as BCWP / ACWP
SPI	Schedule Performance Index	Schedule efficiency of a program. Calculated as BCWP / BCWS
BAC	Budget at Complete	Total allocated budget for a given task

Table 18: Explanatory variables considered in the development of the Ordinary Least Squares models to predict the absolute percent deviation from the EAC and the final ACWP as calculated by the CPI and CR.

Variable Name	Description of Category
Phase	Acquisition phase of the contract: Primarily Development or Production.
Service	Branch/agency in the Government that let the contract: Navy, Air Force, Marine Corps, Army, Department of Defense (Joint), or Missile Defense Agency.
Commodity Type	Majority of product on contract. For example, Aircraft, Missile, Ground Vehicle, Ship, Ordnance, Electronics, etc. A total of 10 types were considered.
Contract Type	Funding outlay as defined in Federal Acquisition Report (FAR) Part 16 Procurement. For example, Fixed-Price Contracts, Cost-Reimbursement Contracts, or Incentive Contracts. A total of 9 types were considered.
MR/TAB	Ratio of Management Reserve to Total Allocated Budget.
Contract Length	Amount of days from the contract’s start date to its completion date.
Money Duration	Amount of days from a contract’s definitized date to its Budget at Completion date.
Delay Length	Amount of days between a contract’s start date and its definitized date.
Max Work Breakdown Structure (WBS)	The maximum level of the WBS for which reporting is indicated for a particular contract.
Percent complete	The percent of completion a contract has currently reached.

Table 19: Inclusion / exclusion criteria for the database for stage one of the analysis with respect to total number of Contract Line Item Numbers (CLINs).

Category	Total CLINs
Initial data extraction from CADE's website	863
CLINs with Over Target Baseline	145
Did not report to at least 99.5% complete	590
Does not meeting initial percent complete requirement of being less than or equal to 20%	15
EAC failed to reach within 5% of the final BAC	17
Final dataset for analysis (Part one)	96

Table 20: Inclusion / exclusion criteria for the database for stages two and three of the analysis with respect to total number of contracts and Contract Line Item Numbers (CLINs).

Criteria	Affected Contracts	Affected CLINs	Total Contracts	Total CLINs
Initial data extraction from CADE's website	451	863	451	863
Experienced an Over Target Baseline	67	145	384	718
Did not report to at least 92.5% complete	163	348	221	370
EAC failed to reach within 10% of the final BAC	24	52	197	318
EAC failed to reach within 5% of the final BAC	29	45	168	273
Reported completion point exceeding 100%	10	19	158	254

Table 21: 95% confidence intervals for the true average completion percentiles for when either the 92.5% cohort or the 100% cohort group contracts achieve two accuracy bands with respect to final contract cost: within 5% and within 10%. Intervals are calculated for both estimates (EAC) determined from CPI or CR. Numbers rounded to one decimal place.

Contract Completion Cohort	Accuracy Metric	Lower Bound	Upper Bound	Sample Size
92.5% Cohort	CPI 5%	70.1	78.2	147
92.5% Cohort	CPI 10%	44.4	54.1	147
92.5% Cohort	CR 5%	74.1	80.4	147
92.5% Cohort	CR 10%	47.7	56.6	147
100% Cohort	CPI 5%	66.9	77.8	107
100% Cohort	CPI 10%	51.0	63.2	107
100% Cohort	CR 5%	71.2	80.0	107
100% Cohort	CR 10%	57.1	67.8	107

Table 22: Hypothesis test results for testing equivalency of the 92.5% and 100% cohort group contracts with respect to being in either the 5% or 10% accuracy band of a contract's final cost. The p-value for the t-tests are based on a two-sided alternative hypothesis with respect to means, while the Wilcoxon rank-sum test is based upon comparing medians. Asterisks denote statistically significant findings at the 0.05 level of significance.

Hypothesis Test	Accuracy Metric	P-value
T-test	CPI 5%	0.5865
Wilcoxon rank-sum test	CPI 5%	0.9064
T-test	CR 5%	0.5277
Wilcoxon rank-sum test	CR 5%	0.7326
T-test	CPI 10%	0.0463*
Wilcoxon rank-sum test	CPI 10%	0.0256*
T-test	CR 10%	0.0036*
Wilcoxon rank-sum test	CR 10%	0.0011*

Table 23: Linear regression analysis results for predicting $AbsCPI_{Dev\%}$. All predictor variables are statistically significant to at least the 0.0001 alpha level. Numbers rounded to two decimal places. Table 9 contains the definitions and general trends of each explanatory variable.

Variable	Parameter Estimate	t-ratio	VIF	Relative % Effect	Cumulative % Effect*
Intercept	38.11	69.19	N/A	N/A	N/A
PC Q4	-33.52	-82.76	1.80	32.80	32.80
PC Q3	-26.46	-55.43	1.56	20.47	53.27
PC Q2	-18.82	-38.10	1.50	13.79	67.06
CPFF	6.11	15.26	1.20	4.95	72.01
WBS 1/2	10.44	13.55	1.08	4.17	76.18
MF Q3	-3.89	-8.56	1.73	3.33	79.51
FPIF	-3.76	-8.09	1.20	2.62	82.13
MF Q1/2	-2.57	-6.18	1.94	2.54	84.67
CL Q1	2.95	7.20	1.41	2.52	87.19
CPAF	4.28	7.64	1.13	2.40	89.59
Electronic	3.05	6.29	1.34	2.15	91.74
AF Aircraft	-3.15	-5.48	1.41	1.92	93.66
Production	-1.79	-5.30	1.26	1.76	95.42
Navy Ship	2.24	4.55	1.51	1.65	97.07
Aircraft	-1.77	-4.26	1.71	1.64	98.71
WBS 11/12	-6.26	-4.28	1.04	1.29	100

Table 24: Linear regression analysis results for predicting $AbsCR_{Dev\%}$. All predictor variables are statistically significant to at least the 0.0001 alpha level. Numbers rounded to two decimal places. Table 9 contains the definitions and general trends of each explanatory variable.

Variable	Parameter Estimate	t-ratio	VIF	Relative % Effect	Cumulative % Effect*
Intercept	35.90	85.46	N/A	N/A	N/A
PC Q4	-35.46	-87.04	1.81	35.33	35.33
PC Q3	-28.29	-59.10	1.56	22.29	57.62
PC Q2	-19.84	-40.04	1.50	14.81	72.43
CPFF	5.99	15.13	1.17	4.93	77.36
WBS 1/2	10.87	14.11	1.04	4.34	81.70
AF Aircraft	-4.36	-8.50	1.11	2.70	84.40
Electronic	3.27	7.37	1.12	2.35	86.75
FPIF	-3.26	-7.19	1.14	2.31	89.06
MF Q4	2.64	6.95	1.21	2.30	91.36
CPAF	3.62	6.55	1.09	2.06	93.42
WBS 11/12	-9.24	-6.39	1.01	1.94	95.36
CL Q1	1.89	4.98	1.21	1.65	97.01
UAV	2.51	5.00	1.14	1.61	98.62
DOD	3.14	4.36	1.10	1.38	100

Table 25: Significant regression variables, their definitions, and general effect on the ability of the EAC_{CPI} or EAC_{CR} to predict a contract's final CAC.

Variable	Definition	Overall affect on CAC Estimate reliability
PC Q4	Contracts in the fourth quartile of percent complete (75% to 100%)	Strongly positive
PC Q3	Contracts in the third quartile of percent complete (50% to 74.99%)	Strongly positive
PC Q2	Contracts in the fourth quartile of percent complete (25% to 49.99%)	Strongly positive
CPFF	Cost-Plus-Fixed-Fee (CPFF) Contract	Moderately negative
WBS 1/2	Contracts with reporting only to Work Breakdown Structure (WBS) level 2 or less	Moderately negative
MF Q1/2	Contracts that have less than 912 days from the definitized date to the BAC date	Minorly positive
MF Q3	Contracts that have between 913 and 1704 days from the definitized date to the BAC date	Minorly negative
MF Q4	Contracts that have between 1705 and 7765 days from the definitized date to the BAC date	Minorly positive
CL Q1	Contracts that have less than 963 days from the contract start date to the completion date	Minorly negative
FPIF	Fixed Price Incentive Fee Target (FPIF) Contract	Minorly positive
CPAF	Cost Plus Award Fee (CPAF) Contract	Minorly negative
WBS 11/12	Contracts with reporting to WBS level 11 or 12	Minorly positive
AF Aircraft	Air Force aircraft contract (manned)	Minorly positive
Aircraft	Aircraft contract (manned) in general	Minorly positive
Electronic	Electronic contract (includes software/avionics)	Minorly negative
UAV	An unmanned aerial vehicle (UAV) contract	Minorly negative
Navy Ship	Navy surface ship contract	Minorly negative
DOD	Contract involving two or more branches of the Department of Defense (DOD)	Minorly negative
Production	A contract solely involving production	Minorly positive

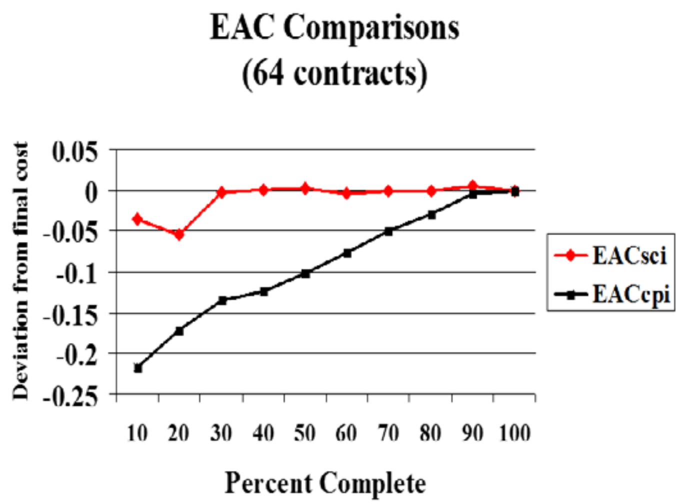


Figure 13: Estimate at Complete (EAC) comparisons graph as shown in Christensen (1996) for EACs calculated from the Critical Ratio (CR) and Cost Performance Index (CPI). Christensen refers to the CR as the Schedule Composite Index (SCI).

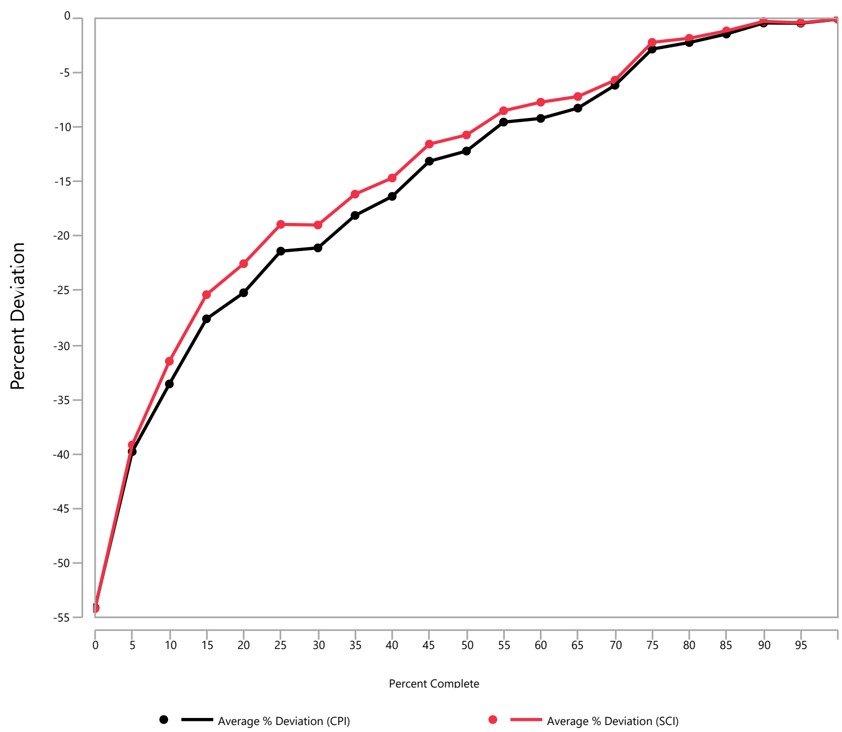


Figure 14: Estimate at Complete (EAC) comparisons graph for EACs calculated from the Critical Ratio (CR) and Cost Performance Index (CPI). Points represent percentile deviation averages of 96 CLINs. In keeping with Christensen (1996), the CR is referred to as the Schedule Composite Index (SCI) on the graph.

V. Conclusion and Recommendations

Overview

This research was initially aimed at replicating Christensen's research in 1996 by using current data to recreate Figure 1. The CPI and SCI are two well-known performance indicators used to estimate EAC. EAC_{CPI} is commonly referred to as an optimistic estimate whereas EAC_{SCI} is known to be a more realistic estimate of the final cost. According to Christensen's study in 1996, contracts showed stability earlier when SCI was used as the performance indicator when estimating EAC. When Christensen analyzed his dataset using 64 contracts, he concluded that EAC_{SCI} is a more accurate indicator of the final cost at complete; when EAC_{SCI} was used to predict the final CAC, the deviation from the final cost was less than 5% starting at the 20% complete point. When EAC_{CPI} was used to predict the final cost at complete, the deviation from the final cost was less than 5% at the 70% complete point. This indicated that EAC_{SCI} could help make more informed decisions on programs and ultimately save money by making wiser decisions on whether to continue or stop funding certain contracts.

However, the results from this research show that neither EAC_{CPI} nor EAC_{SCI} are accurate predictors of the final cost at complete until the 70% complete point. Although EAC_{SCI} is still a more accurate indicator of the final cost at complete according to Figure 7, the 5% accuracy range is not reached until the 70% complete point which leads one to suspect the accuracies of reported CPI and SPI metrics.

After having taken the averages at each percent complete to produce Figure 7, the absolute values of $\%DCAC_{CPI}$ and $\%DCAC_{SCI}$ were taken to negate the influence of

positive and negative numbers for Parts II and III of the analysis. Part II re-analyzed Tracy and White's (2011) work to test if there was a statistical difference in %DCAC between contracts that were 92.5% complete versus those that were 100% complete. There was a statistical difference between the two groups for the 10% threshold which indicates that a substantial amount of ACWP is reported towards the end of a contract. Next, a population analysis was used to determine the main drivers of $|\%DCAC_{CPI}|$ and $|\%DCAC_{SCI}|$; the main drivers between the two models were identical to one another. The models indicate that the main drivers of |%DCAC| are the percent complete. The moderate drivers are CPFF and FPIF contracts, Navy surface ship contracts, contracts with WBS level 1 or 2, and electronic/software contracts.

Research Questions Answered

The three research questions from Chapter 1 are answered below.

1. Which efficiency factor is most accurate at predicting CAC?

Figure 7 displays the final graph from the research. According to the graph, using SCI as the performance index is still a more accurate predictor of CAC but not by a significant amount. When comparing Figures 1 and 7, EAC_{CPI}'s are relatively comparable with respect to when they achieve the within 10% threshold. In 1996, the percentage was approximately 50% while modern contracts suggest this percentage is closer to 55%. In 1996, the percent complete for when EAC_{SCI} was within 10% of CAC was essentially at contract initiation. For modern contracts, it appears EAC_{SCI} achieves this 10% accuracy band at approximately 50% completion.

2. At what percent complete does the EAC get within 5% of the CAC?

Both EAC_{CPI} and EAC_{SCI} achieve the 5% accuracy band at approximately 70% completion with EAC_{SCI} achieving this threshold slightly quicker according to Figure 7.

3. What are the major and moderate drivers that influence |%DCAC|?

The model indicates that that the main drivers of |%DCAC| are the percent complete. Contracts in the first quartile of percent complete have a strongly negative effect on CAC estimate reliability while contracts in the third and fourth quartiles of percent complete have a strongly positive effect on CAC estimate reliability. This makes sense because estimates become substantially more accurate as contracts gets closer to the end stages. The variables that had a moderately negative effect on CAC, in reference to Table 15, are CPFF contracts, Navy surface ship contracts (Ship), contracts with WBS level 1 or 2 (WBS 1,2), and Electronic/Automated software contracts. FPIF contracts had a moderately positive effect on CAC.

Recommendations

We recommend that program offices no longer use EAC_{SCI} as an upper estimate for CAC. EAC_{SCI} is, at the very most, a supplementary lower estimate that should be used instead of EAC_{CPI}. Accordingly, program offices should use a different approach when estimating CAC instead of using the two standard approaches of using CPI and SCI to calculate EAC. Program managers should view CPI and SPI with caution as these

numbers did not produce reliable EACs. The research indicates that since EAC_{CPI} and EAC_{SCI} are so similar, the reported SPI is close to 1.

Recommendations for Future Research

Our research was conducted on CLINs and relied on the accuracy of data provided by CADE. For future research, we recommend delving into the reporting procedures of CPI and SPI. Because the CPIs and SPIs provided for this research were high numbers, this resulted in low estimates for EAC. Also, we recommend looking at Level of Effort (LOE) to see if this is one of the main drivers of |%DCAC|; LOEs are man hours associated with CLINs.

References

Anbari, F.T. (2003), “Earned Value Project Management Method and Extensions”, *Project Management Journal*, Vol. 34 No. 4, pp. 12-23.

Bollen, K.A. and Jackman, R.W. (1990), “Regression diagnostics: an expository treatment of outliers and influential cases”, in Fox, J. and Long, J.S. (Eds.), *Modern Methods of Data Analysis*, Sage, Newbury Park, CA, pp. 257-91.

Christensen, D.S. (1996), “Project advocacy and the estimate at completion problem”, *The Journal of Cost Analysis*, Vol. 13 No. 1, pp. 35-60. doi:10.1080/08823871.1996.10462298

Christensen, D.S. and Payne, K.I. (1992), “Cost Performance Index Stability – Fact or Fiction?”, *Journal of Parametrics*, Vol.10, pp. 27-40. doi:10.1080/10157891.1992.10462509

Defense Acquisition University (2017), “Earned Value Management ‘Gold Card’”, available at: https://www.dau.mil/tools/Lists/DAUTools/Attachments/19/20170720-gold%20card%201%20sided.pdf (accessed 3 February 2018).

Defense Contract Management Agency (2006), *Earned Value Management Implementation Guide (EVMIG)*, Author, Alexandria, VA.

Defense Contract Management Agency (2012), *Earned Value Management System (EVMS) Program Analysis Pamphlet (PAP) DCMA-EA PAM 200.1*, Humphreys Associates, Wright Patterson AFB, OH.

Department of the Air Force (2007), *Air Force Cost Analysis Handbook*, Author, Washington, DC.

Department of Defense (DoD) (2008), *Operation of the Defense Acquisition System* DoD Instruction 5000.02, Author, Washington, DC.

Earned Value Management (2017), “ANSI EIA 748 Earned Value Management”, available at: http://acqnotes.com/acqnote/tasks/ansi-eia-748-earned-value-management (accessed 12 February 2018).

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

Kranz, G.M., and Bliss, G.R. (2015), *Department of Defense Earned Value Management System Interpretation Guide OUSD AT&L (PARCA)*, U.S. Government Printing Office, Washington, DC.

Lessard, C. & Lessard, J. (2007), Project Management for Engineering Design, Morgan & Claypool, San Rafael, CA.

Lipke, W. (2003), "Schedule is Different", *The Measurable News*, March & Summer 2003

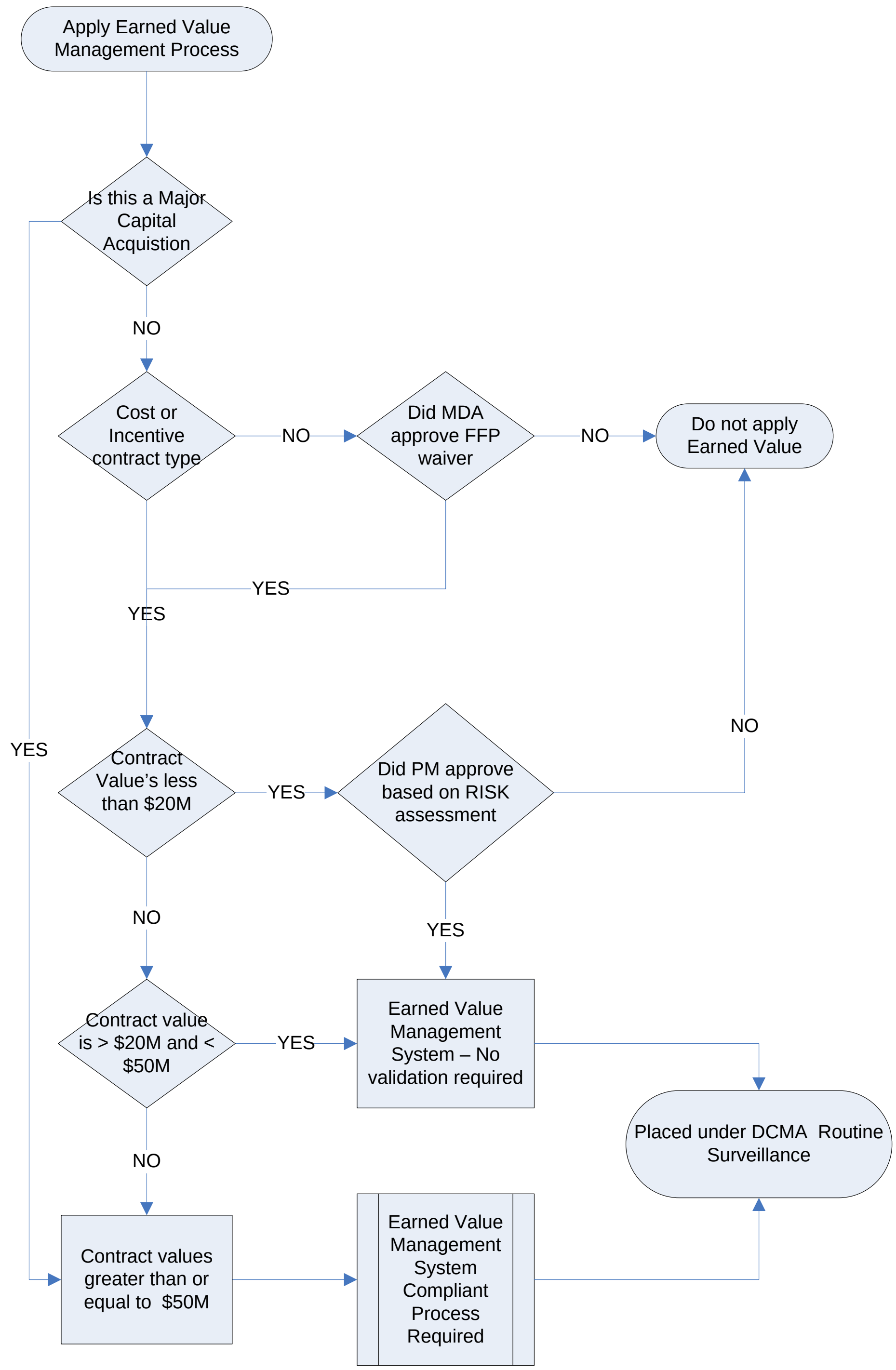
Office of the Under Secretary of Defense: Acquisition, Technology, and Logistics (USD-AT&L) (2008), Office of the Secretary of Defense Central Repository for Earned Value Management Data, U.S. Government Printing Office, Washington, D.C.

Petter, J.L., Ritschel, J.D. and White, E.D. (2015), "Stability properties in Department of Defense contracts: answering the controversy", *Journal of Public Procurement*, Vol. 15 No. 3, pp. 341-364.

Rose, J.M. (2002), "Earned Value Management (EVM) [PowerPoint slides]", available at: http://documentcenter.meetingcaddie.mobi/Events/1928d4fe-0359-420f-a3e3-12ae00fb16ab/Documents/ad339ff0-ad12-4b8b-b45d-36a2f5d1cb19.pdf

Tracy, S.P. and White, E.D. (2011), “Estimating the final cost of a DoD acquisition contract”, *Journal of Public Procurement*, Vol. 11 No. 2, pp. 190-205.

Appendix A: EVM Compliance Chart (EVMIG, 2006)



Appendix B: EVM Criteria (NDIA, 2005)

The following EVM criteria are from the National Defense Industrial Association (NDIA) Program Management Systems Committee (PMSC) American National Standards Institute/Electronic Industries Alliance (ANSI/EIA) Standard 748, Earned Value Management System.

- Criterion 1. Define the authorized work elements for the agency. A WBS, tailored for effective internal management control, is commonly used in this process.
- Criterion 2. Identify the organizational structure including the major contractors responsible for accomplishing the authorized work, and define the organizational elements in which work will be planned and controlled.
- Criterion 3. Provide for the integration of the agency's planning, scheduling, budgeting, work authorization and cost accumulation processes with each other, the WBS, and the OBS.
- Criterion 4. Identify the organization or function responsible for controlling overhead (indirect costs).
- Criterion 5. Provide for integration of the WBS and the organizational structure in a manner that permits cost and schedule performance measurement by elements of either or both structures as needed.
- Criterion 6. Schedule the authorized work in a manner that describes the sequence of work and identifies significant task interdependencies required to meet all authorized requirements.

- Criterion 7. Identify physical products, milestones, technical performance goals, or other indicators that will be used to measure progress.
- Criterion 8. Establish and maintain a time-phased budget baseline, at the control account level, against which performance can be measured. Budget for far-term efforts may be held in higher-level accounts until an appropriate time for allocation at the control account level. Initial budgets established for performance measurement will be based on either internal management goals or the external customer- negotiated target cost including estimates for authorized but undefined work.
- Criterion 9. Establish budgets for authorized work with identification of significant cost elements (labor, material, etc.) as needed for internal management and for control of contractors.
- Criterion 10. To the extent it is practical to identify the authorized work in discrete work packages, establish budgets for this work in terms of dollars, hours, or other measurable units. Where the entire control account is not subdivided into work packages, identify the far term effort in larger planning packages for budget and scheduling purposes.
- Criterion 11. Provide that the sum of all work package budgets plus planning package budgets within a control account equals the control account budget.
- Criterion 12. Identify and control level of effort activity by time-phased budgets established for this purpose. Only that effort which is immeasurable or for which measurement is impractical may be classified as level of effort.

- Criterion 13. Establish overhead budgets for each significant organizational component of the company for expenses that will become indirect costs. Reflect in the budgets, at the appropriate level, the amounts in overhead pools that are planned to be allocated as indirect costs.
- Criterion 14. Identify management reserves and undistributed budget.
- Criterion 15. Provide that the allocated budget is reconciled with the sum of all internal budgets and management reserves.
- Criterion 16. Record direct costs in a manner consistent with the budgets in a formal system controlled by the general books of account.
- Criterion 17. Summarize direct costs from control accounts into the WBS without allocation of a single control account to two or more WBS elements.
- Criterion 18. Summarize direct costs from the control accounts into the agency's organizational elements without allocation of a single control account to two or more organizational elements.
- Criterion 19. Record all indirect costs that will be allocated to the agency.
- Criterion 20. Identify unit costs, equivalent units costs, or lot costs when needed.
- Criterion 21. For EVMS, the material accounting system will provide for:

  - Accurate cost accumulation and assignment of costs to control accounts in a manner consistent with the budgets using recognized, acceptable, costing techniques.

  - Cost performance measurement at the point in time most suitable for the

category of material involved, but no earlier than the time of progress payments or actual receipt of material.

- Full accountability of all material purchased including the residual inventory.

- Criterion 22. At least on a monthly basis, generate the following information at the control account and other levels as necessary for management control using actual cost data from, or reconcilable with, the accounting system:

  - Comparison of the amount of planned budget and the amount of budget earned for work accomplished. This comparison provides the schedule variance. –

  - Comparison of the amount of the budget earned with the actual (applied where appropriate) direct costs for the same work. This comparison provides the cost variance.

- Criterion 23. Identify, at least monthly, the EV Variance between both planned and actual schedule performance and planned and actual cost performance, and provide the reasons for the variances in the detail needed by management.

- Criterion 24. Identify budgeted and applied (or actual) indirect costs at the level and frequency needed by management for effective control, along with the reasons for any significant variances.

- Criterion 25. Summarize the data elements and associated variances through the organization and/or WBS to support management needs and any customer reporting specified in the contract.

- Criterion 26. Implement managerial actions taken as the result of earned value

- Criterion 27. Develop revised EAC based on performance to date, commitment values for material, and estimates of future conditions. Compare this information with the performance measurement baseline to identify variances at completion important to company management and any applicable customer reporting requirements including statements of funding requirements.
- Criterion 28. Incorporate authorized changes in a timely manner, recording the effects of such changes in budgets and schedules. In the directed effort prior to negotiation of a change, base such revisions on the amount estimated and budgeted to the organizations.
- Criterion 29. Reconcile current budgets to prior budgets in terms of changes to the authorized work and internal replanning in the detail needed by management for effective control.
- Criterion 30. Control retroactive changes to records pertaining to work performed that would change previously reported amounts for actual costs, earned value, or budgets. Adjustments should be made only for correction of errors, routine accounting adjustments, effects of customer or management directed changes, or to improve the baseline integrity and accuracy of performance measurement data.
- Criterion 31. Prevent revisions to the agency budget except for authorized changes.
- Criterion 32. Document changes to the performance measurement baseline.

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 074-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
23-03-2018	Master's Thesis	October 2016 – March 2018

TITLE AND SUBTITLE	
An Analysis of the Estimate at Complete for Department of Defense Contracts	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Kim, Deborah B., First Lieutenant, USAF	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/ENC) 2950 Hobson Way, Building 640 WPAFB OH 45433-8865	AFIT-ENC-MS-18-M-214
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)** **AFIT/LSB**
Air Force Institute of Technology School of Systems and Logistics 2950 Hobson Way WPAFB, OH 45433-8865 ATTN: Chad Millette (Chad.Millette@afit.edu)	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRUBTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
When contractors provide timely and reliable information on the status of a contract, both contractors and government program offices can provide an accurate estimate of a contract's completion costs. This research shows that the cumulative cost performance indices provided by contractors and program offices are high and less accurate than those of previous years and/or that a significant amount of ACWP is being documented in the final portion of a contract. The high performance indices resulted in EACs that were low-balled during the majority of a contract's life which shows a need to improve the use of EVM metrics for effective project and program management. This research replicates Christensen's findings in 1996 which proved that using the SCI was a more accurate indicator of EAC vice the CPI. Consistent with Christensen's research, SCI is still a more accurate indicator of CAC according to this study but not by a significant amount. When SCI is used to predict the final cost on contracts from the 21st century, a 5% deviation from the final cost starts at the 70% complete point when using both CPI and SCI as the performance indicator.

15. SUBJECT TERMS
Earned Value Management (EVM), Estimate at Completion (EAC), Percent Complete, EAC Stability

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Edward D. White, AFIT/ENC
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	105	(937) 255-3636 x4540(Comm) (Edward.White@afit.edu)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18