3-23-2017

Using Multiple and Logistic Regression to Estimate the Median Will-Cost and Probability of Cost and Schedule Overrun for Program Managers

Ryan C. Trudelle

Using Multiple and Logistic Regression to Estimate the Median Will-Cost and Probability of Cost and Schedule Overrun for Program Managers

THESIS

Ryan C. Trudelle

Captain, USAF

AFIT-ENC-MS-17-M-231

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENC-MS-17-M-231

Using Multiple and Logistic Regression to Estimate the Median Will-Cost and Probability of Cost and Schedule Overrun for Program Managers

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

Ryan C. Trudelle, B.S.

Captain, USAF

March 2017

AFIT-ENC-MS-17-M-231

Using Multiple and Logistic Regression to Estimate the Median Will-Cost and Probability of Cost and Schedule Overrun for Program Managers

Ryan C. Trudelle, B.S.

Captain, USAF

Committee Membership:

Edward D. White, Ph.D
Chair

Lt Col Brandon M. Lucas, Ph.D
Member

Lt Col Clay M. Koschnick, Ph.D
Member

Lt Col Jonathan D. Ritschel, Ph.D
Member

AFIT-ENC-MS-17-M-231

Abstract

The main concern of a program manager is to manage the cost, schedule, and performance triad of a program. Historically, programs tend to meet the performance aspect at the expense of cost or schedule, or both. This research gives the acquisition community a set of tools that enables them to impartially analyze the cost and schedule of their programs, helping to mitigate these issues. Five regression models encompass this toolset; one to estimate the median program cost and four to identify the probability of realizing a given overrun. The cost model explains 81% of the variation in program acquisition using seven predictor variables available to the estimator at the time of MS B start. Four logistic models estimate the probabilities that a program may identify as a program that experiences cost and schedule overruns of specific magnitudes from their MS B estimate. These models predict the group the program may reside in with an accuracy of at least 0.79 probability and use multiple predictor variables available at MS B. With these tools the program manager has the ability to preemptively identify potential problems in their program based on the program's characteristics, potentially saving millions in cost and schedule overruns.

Acknowledgments

First and foremost, I want to thank my wife, who knows more than anyone else involved in this process, just how much this thesis took from me and how much I poured into it. Through the late nights spent compiling data, often past 2AM and even up to 5AM with no questions asked until I awoke after noon, she supported me and cared for our beautiful children. Thank you for being my rock, the one relative constant I could return to on a daily basis. Thank you also to my two children, who were a constant reminder of what I was working so hard for. My daughter's constant enthusiasm for going to school and inquisition into what daddy did at school provided a welcome grin on a daily basis while my son's calm demeanor and innocent smile became a rallying point for me when times seemed overwhelming.

A special thanks goes out to my advisor Dr. White who knows the sheer amount of work and changes that went into this thesis in order to churn out a final polished product. Without you, I fear I might still be lost in the weeds, so thank you for helping me see the big picture as well as being my sounding board for ideas and keeping my crazy writing in check. Only we will ever know the difference between what I gave you and what the output was, but I can say the difference is great.

Thank you to my committee, who gave a damn enough to put in many, many hours of their own time to lend their expertise in the realm of cost and schedule. Each of you were selected for a purpose and you fulfilled your roles expertly. Without their help I would have been hard pressed to form a coherent and competent analysis regarding the acquisition process. An additional thanks to Mrs. Turner, without you everyone would know exactly how horrendous my grammar truly is; thanks so much for the help.

Table of Contents

List of Figures

List of Tables

Using Multiple and Logistic Regression to Estimate the Median Will-Cost and Probability of Cost and Schedule Overrun for Program Managers

I. Introduction

General Issue

The intersection of schedule and cost is of paramount importance, as it makes up two of the three parts of the Acquisition Program Baseline (APB)—the third being performance. Managing both cost and schedule to achieve the desired performance is what a program manager does by definition. Establishing realistic baselines for schedule and cost are exceedingly difficult due to the nature of the Department of Defense (DoD) acquisition environment, as well as the levels of risk and uncertainty due to the nature of the program itself.

To combat the risk and uncertainty in cost and schedule, a program may err on the side of caution—lengthening schedules and increasing cost estimates to unreasonably high levels in order to capture all variability. This strategy is dangerous to the program and the community as a whole in that it not only delays the fielding of an identified need to the warfighting community; it also needlessly ties up funding that could be used to fund other critical assets.

On the opposite end of the spectrum lie the programs that are overly optimistic in terms of cost and schedule. This strategy is equally dangerous in terms of potential negative outcomes to both schedule and cost. A schedule that is overly optimistic runs the risk of experiencing unscheduled delays due to potential re-work, if a problem is discovered that could have been prevented, and the schedule allowed for the proper oversight. The expenses of conducting this re-work can increase at an exponential rate as the program moves through the acquisition process as demonstrated by Figure 1. Being overly optimistic in terms of program cost can lead decision makers to take critical funding from other programs or force a program manager to delay the program until additional funding can be secured.

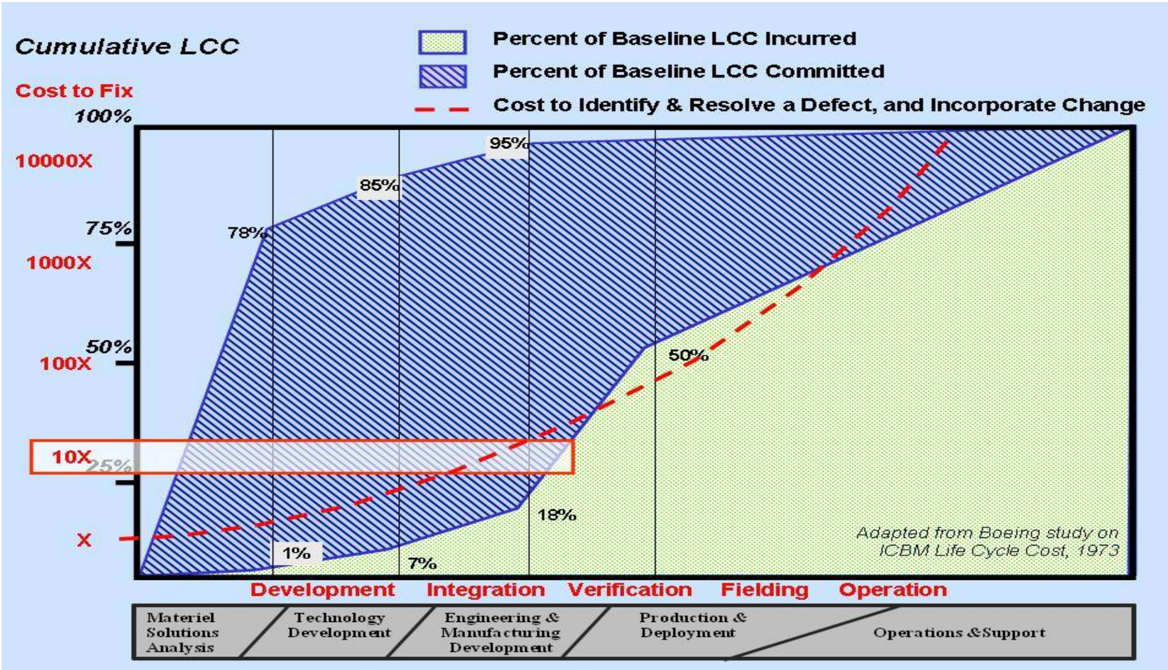

Figure 1: Cumulative LCC over Program Life

Specific Issue

There is a great deal of research that has been conducted concerning quantitatively optimizing and predicting cost in a DoD program environment. Research conducted by Brown, White, Ritschel, and Siebel (2015) identified cost expenditure trends and growth patterns. Birchler, Christle, and Groo (2011) researched concurrency's effect on cost growth using multiple regression. There are many more examples of cost research and they are explored in detail in Chapter II of this thesis.

One research project conducted at the Air Force Institute of Technology (AFIT) developed a quantitative method to predict a program schedule based on specific program attributes. Prior to this method, and up to this point, the Air Force standard for estimating a program schedule has been to rely on the opinion of a subject matter expert (SME), or to rely on the actual observed schedule of a similar completed program and adjust for differences (Jimenezet al., 2016).

There has been no past research at AFIT that combines both cost and schedule into quantitative models within a set of programs to provide benchmarks for program evaluation. The goal of this research is to provide the Air Force acquisition community with a set of tools to analyze where they may realize cost or schedule savings. At the very least, we strive toward an unbiased method of generating a Will Cost estimate to serve as a benchmark to compare to the actual program estimate and to identify what program characteristics may lead to significant or critical overruns to the current program baseline. These, in turn, give program management a starting point to look at reducing cost and schedule further as per the Should Cost and Should Schedule directives.

Scope and Limitations of Research

The scope of this research is constrained to predicting a benchmark for program cost starting at program initiation, Milestone-B (MS-B), through to Initial Operational Capability (IOC) and for estimating the likelihood a program may stay within certain cost and schedule bounds. The Defense Acquisition University (DAU), a government organization created to provide the DoD acquisition personnel with a professional career path and training, defines IOC thusly: “In general, attained when selected units and/or organizations in the force structure scheduled to receive a new system have received it and have the ability to employ and maintain it” (Defense Acquisition University, 2016). The official definition for IOC continues by stating that each particular program defines its own IOC. With each program studied having its own individual definition for IOC, there is a great deal of known variability in any model that utilizes IOC as an evaluation point. Removing this variability by researching and determining a universal definition for what IOC could be yields a more reliable model for predicting program schedule. This research is limited further by only including programs that have cost data from the Technology Development phase (pre-MS-B), since this data is used as a predictor variable in the schedule model (Jimenez, 2016).

The definition we use as IOC is a limitation within the research due to the unavailability of several data points concerning IOC within the dataset, as well as the highly subjective nature of the definition of when IOC occurs in a program. Due to these limitations, research alternatives to IOC as a termination point for our research. Access to the information that makes it possible to create this new definition is also a limitation, as it is not readily available.

We utilize Selected Acquisition Report (SAR) to populate the data within the dataset being analyzed. The SAR database contains the reported financial and schedule information for select programs (Brown, White, Ritschel, & Seibel, 2015). The initial database from the SAR data is compiled and available for us to utilize, and additional data points have been found using SARs as well as other historical program documents. Another limitation we note is that only programs that can be included in the regression model we use to predict schedule are available for inclusion in the analysis for the cost model. We acknowledge that this approach may not be the optimal method; however, due to the constraints on the data and the prior research, we are confident moving forward by limiting the amount of programs to only those that fit within the schedule model parameters.

Research Objectives

There are two main objectives of this research: to develop a statistical model to output the median estimate for program Will-Cost, and to develop logistic regression models to estimate the probability a program will overrun certain threshold values pertaining to both cost and schedule. The Will-Cost multiple-regression model is intended to give the acquisition community an unbiased median cost estimate for a program, which can, in turn, be used to develop the Should-Cost estimate, in millions of dollars for BY17, for said program. The logistic regression models for the overrun analysis give the decision makers in the Air Force community a meaningful analysis for what program characteristics are more likely to experience cost and schedule overruns of the magnitudes identified in this study.

Research Questions

Our research is focused on addressing two questions. Initially we address the question, "How can we use and build upon a previously created database to develop a mathematical model to predict the median cost of a program?" The second question we address is: "How can we identify program characteristics for significantly and critically overrunning either cost, schedule, or both given the current APB, at MS B through IOC, and predict the probability that a program will experience such overruns?"

Summary

Creating an objective and statistically-sound method to predict the median program cost provides the acquisition community with a tool to impartially estimate their Will-Cost, which allows them to take a unique perspective on potential program efficiencies to affect their Should-Cost goals. By analyzing the program cost and schedule estimates from MS B and comparing them to the actuals based on the thresholds for significant and critical program overruns we are able to give the program management community tools to identify potential problem programs and allocate resources accordingly.

A review of literature is conducted in Chapter II to identify potential program stabilization points as a substitute for IOC, as well as reasons for cost and schedule variance and their associated predictors. In Chapter III, the literature review is used to lay the foundation for our database and data utilization methods. In Chapter IV, we build, test, and validate a multiple-regression model for estimating the program median Will-Cost. In Chapter V we analyze the program indicators for not meeting their cost or schedule estimates and calculate the probabilities for programs, given certain characteristics, to fall into one of three groups. In Chapter VI, we close with the conclusions of our research and any potential follow-on research.

II. Literature Review

Chapter Overview

In this chapter, we conduct a review of the literature pertaining to relevant research into DoD cost and schedule analysis in order to build a foundation and gain the insight required to justify and conduct our research. In order to gain this insight, we provide an overview of the research conducted in the past and highlight how we utilize what they found to move forward with our research. In addition to identifying relevant research, we identify where we add additional insights by conducting our research; specifically, we identify where we may alleviate some of the research deficiencies.

The ensuing sections provide the background for building our database, validating the previously built schedule benchmark model, building a cost benchmark model within the scope of our schedule model, and combining the results from each model to gain insight into how schedule and cost interact. Before cost and schedule are addressed, we take time to explore the issue of using IOC as the conclusion point of our models.

Defining IOC

The focus of this thesis is to develop a model to estimate the median cost of a DoD acquisition program and to identify the factors that may lead to a program overrunning certain cost and schedule thresholds. However, IOC is an integral piece in achieving both of these goals. Therefore, the literature review begins by researching what IOC is, the issues that we have with using it in this form, and how it may affect the research. It is important to note that we have not found any peer-reviewed research defining IOC into a more rigid concept.

IOC is a program milestone that is set by the program's user community and program manager (PM) following the guidelines set by the acquisition community. DAU defines IOC as being "attained when selected units and/or organizations in the force structure scheduled to receive a new system have received it and have the ability to employ and maintain it." ("Defense Acquisition University (DAU) 'IOC Definition,'") . The definition purposefully leaves out any stringent empirical requirements for meeting IOC and standardized the programs only by requiring all IOC dates to meet the minimum operational capabilities of the user community; this simply means that whatever they are acquiring, it has to work. This subjective treatment of IOC creates difficulty when objectively analyzing the data using a multiple-regression approach.

By having an end date to our research that varies from program to program in a non-uniform manner, we have introduced a great deal of known variability. When developing a multiple regression model, one main objective is to explain as much of the variability as possible—this makes the prediction more valid and accurate since the model can account for more of the variability that is present within the population being studied. By accounting for this source of known variability within our models, we are much more precise in predicting where a program should be benchmarked for cost and schedule.

Past research has successfully used IOC as the end date, because it is generally outlined in the SAR and readily available as a data point for each program. Research has even found that approximately 91% of cost growth, experienced on certain aircraft and avionics programs, is realized by the date each program defines as IOC (Kozlak, 2016). This fact allows us to continue utilizing the given programmatic IOC dates in the model despite the risk to increased variability. It is clear to us that a universal definition placing the finish line for each program in the same place, whether that be a specific percent of total units fielded or the first unit equipped, would be a better marker for reducing the variability when comparing multiple programs.

Moving forward, we explore how cost and schedule have been analyzed in the past.

Schedule Research

A program's schedule defines many aspects of the program: when the funding needs to be spent, how much it costs, and when it delivers a critical requirement to the warfighter, to name a few. Schedule overruns have been, and continue to be, a major issue in defense acquisition programs. To combat this issue, the Secretary of the Air Force, Deborah James, introduced a new initiative, called "Should Schedule", in 2015. Should Schedule created a benchmark for how long a program's schedule should be, if identified process inefficiencies and risks were driven out of the program (James, 2015). Essentially, should schedule develops a goal for which a program should strive towards.

In the wake of this new Should Schedule requirement, which was levied on all acquisition category I (ACAT I) programs, the largest and most highly visible programs in the Air Force (AF), research conducted at AFIT developed a multiple regression model which could be used to predict how long a program's schedule should be, from MS-B to IOC, based on historical figures. This model is successful at predicting a realistic program schedule—explaining 42.9% of variability (Jimenez et al., 2016). The research was based on the inclusion of 56 programs and relied upon three separate predictor variables that were deemed statistically significant. The predictor variables were whether a program was a new start or a modification to an existing platform, the year MS-B occurred as it related to a specific change in defense acquisition policy, and the amount of funding prior to MS-B (Jimenez et al., 2016). This tool has already been used, in an unofficial capacity, in the field by the KC-46 tanker program as a "cross check" for the latest schedule estimate. The tool produced an estimate that was within four months of the latest revised program schedule (E-mail in Appendix C).

In 2015, Brown, White, Ritschel, and Seibel conducted research to determine if there is a better, more empirical method to estimate a programs expenditure curve over time. The widely accepted method utilized in the acquisition community is the 60/40 “rule of thumb”, which states that at a program’s halfway point, roughly 60% of the costs are accrued and 40% are remaining. Specifically tying their research to aircraft development programs, they tested the 60/40 heuristic and determined that it could not account for the differences between new start and modification programs. Next, they researched program characteristics to construct an aircraft-centric methodology. Lastly, they compared the accuracy of their models to the baseline 60/40 model—finding that their Weibull model explained 74.6% of total variation, a 6.5% increase over the 60/40 heuristic (Brown et al., 2015).

The most important aspect of Brown’s research is utilized in our own research, which was used in Jimenez’s research as well. There is a statistically significant shift in program obligation expenditure rate between programs that began development prior to 1985 and those that began post-1985 (Brown et al., 2015). This shift in expenditure rate is believed to be caused by President Reagan’s Blue Ribbon Commission on Defense Management (also referred to as the Packard Commission), which was aimed at studying issues within the DoD acquisition process (Packard, 1986). From this report came several recommendations, which are credited with having led to many subsequent acquisition reform laws. These laws, in turn, affected the nature of every acquisition program within the DoD since 1985. Given this knowledge, we evaluate a binary variable which accounts for programs starting either before or after 1985, which Jimenez et al. (2016) found highly predictive of schedule. Without this variable, we would not be evaluating programs on an equal basis since they are subject to different rules.

Monaco and White (2005) attempted to predict schedule risk using a regression approach.

They modified and used the database built by Sipple et al. (2004), and subsequently modified by future research as mentioned in the “A Benchmark for Cost” section of this chapter, to include 67 programs in total. The goal of their research was to predict the probability, and associated magnitude, of schedule growth within the programs studied (Monaco and White, 2005). Of particular note within this research is the observation of data availability issues within the SAR database.

Monaco expanded the research database to include programs from 1990 to 2003 that have completed the Engineering and Manufacturing Development (EMD) phase of acquisitions using the final SAR report to populate their data. They built a logistic regression model to predict the probability of schedule growth occurring and a multiple regression model to predict the magnitude of schedule growth, given growth has occurred. When building their database to perform these analyses they note that several potentially highly predictive variables cannot be used due to their unavailability within the SARs. Most notably, the variable missing from the majority of programs was the IOC date from either MS-B or EMD contract award (Monaco, 2005). Jimenez et al. (2016) note similar difficulties in obtaining an IOC date using the SAR database. Despite the issues in obtaining data points for all of the identified potential predictors of schedule, Monaco was able to build two highly predictive models.

Gailey III (2002) expanded upon the research conducted by Reig (1995) and increased the database studied from 24 to 46 programs and re-evaluated the results. The study focused on 28 program characteristics that were analyzed to determine if any of them could be a predictor of performance during the engineering and manufacturing development (EMD) phase. In order to do this, they split the programs into two groups using EMD duration overrun as the criterion and defining the group with the larger overrun as “bad”, or unsuccessful (Gailey III, 2002).

Cost Research

The cost of a program is of paramount importance and feeds into every aspect of the program—from schedule to performance. Decision makers want to know where to allocate precious resources and, to do this, they routinely rely on point estimates from the cost community that may not be bound by historical facts or are overstated for budgetary concerns. To combat this issue, the USD(AT&L) issued a memorandum in 2011, which mandated the use of Will-Cost and Should-Cost management for all programs meeting the requirements. The Will-Cost estimate is what the current program estimate is, which is to be compared to the Should-Cost estimate. The Should-Cost estimate is meant to be what a program should cost if all process inefficiencies are driven out; essentially, this becomes a benchmark for the program to strive towards (MFR in Appendix A). There have been several empirical models aimed at predicting program cost, none of which are widely used across the DoD. We use the completed research to vector our own research in order to develop a complete and accurate model.

The most recent cost research conducted at AFIT, completed in 2016, focused on cost growth amongst the programs studied. More specifically, the research is focused on identifying factors that contribute to cost growth and what the historical cost growth is at the four major program reviews. Most importantly to our research, they identified that bombers, prototyping, and electronic aircraft systems upgrades were the most common predictors of cost growth at these program reviews and that at IOC, the median percent of total cost growth experienced is 91% (Kozlak, 2016). We use this information to identify and evaluate these three predictor variables for inclusion into our cost-estimating model. The findings that IOC was a point in the program lifecycle, where the majority of cost growth was realized, means that our assumptions to continue using IOC as a program termination date is valid. This is critical since our schedule model does not include any programs for which we cannot ascertain the IOC date and cost information.

The study conducted by Deitz, Eveleigh, Holzer, & Sarkani (2013) examined the importance of developing a robust Analysis of Alternatives (AoA) prior to MS-B and the effects that an analysis of this nature may have on program success. According to their research, the DoD has a history of rushing programs into development that are simply not ready due to various reasons. In the case of the Joint Strike Fighter, the Government Accountability Office (GAO) attributed a major factor for the cost overrun to the program prematurely entering into EMD. The researchers stated that, while only 10% of a program's life-cycle cost was invested prior to MS-B, it may be the most important 10% because 70% of a program's life-cycle costs are committed and set at this phase (Deitz et al. 2013). These findings indicate that the most important time for cost savings in a program is prior to MS-B and are important for our research because they validate using pre-MS-B data in our estimating models.

Birchler, Christle, & Groo (2011) studied the idea of concurrency, or developing a weapons system while in production, and cost growth. The researchers acknowledged that concurrency could inherently increase program risk, which could foreseeably lead to cost growth, as is the consensus in the acquisition community. They made the point that every program had some level of concurrency, and concurrency is necessary in any environment other than a zero-risk environment. Their research is focused on exploring the relationship between concurrency and cost growth (Birchler et al., 2011).

In their study, they defined concurrency as "the proportion of research, development, test and evaluation (RDT&E) appropriations that are authorized during the same years that procurement appropriations are authorized" (Birchler et al., 2011:5). They further defined concurrency in their research as the point where less than 95% of the total amount of RDT&E funding is still being spent, while procurement is taking place. This metric was used because they reason that RDT&E funding is spent throughout most of a program's life for various reasons that do not necessarily deal with concurrency and, therefore, anything less than 5% of the total amount of RDT&E funding being spent at this time did not constitute concurrency. The researchers used multiple regression techniques to determine if concurrency predicted cost growth and found no evidence to support the relationship (Birchler et al., 2011).

Although Birchler et al. (2011) found no evidence that concurrency was correlated with cost growth, their database was much smaller than our database—consisting of only 28 programs. This leads us to believe they may not have had enough power to detect correlation. Additionally, we can evaluate concurrency as a possible predictor for schedule, even though Jimenez's research did not find that it was highly predictive of schedule, as we also expand this database and analyze more programs than Jimenez.

Foreman (2007) researched methods to further cost and schedule growth estimation by including longitudinal variables that account for changes that take place over time. The research conducted by Foreman et al. uses previous research conducted at AFIT in the early 2000's by Sipple et al. (2004), Genest & White (2005), Lucas (2004), and others by utilizing the database built and added to by these researchers and analyzing additional predictor variables. Most importantly, the researchers found that the length between MS-III (synonymous with MS-C) and IOC, as well as a binary variable for MS-III slip (indicating whether or not the MS-III date occurred later than originally planned), to be predictive and positively correlated to cost growth in a program (Foreman, 2007). These indicate to us that schedule overruns, particularly later in a program, may correlate to higher costs.

Sipple et al. (2004) conducted a study to explore defense acquisition program cost growth using SAR data from 1990 to 2000 for programs from all defense departments. The focus of this research is on cost growth as it applies to the Engineering Manufacturing Development (EMD) phase, or between MS-B and MS-C, of an acquisition program. This research is further limited by focusing solely on engineering cost growth within this phase. Applying a multiple regression analysis to the programs that experience cost growth within their dataset, Sipple et al. developed a model that predicts the expected amount of cost growth with an adjusted R^2 value of .4645. It is important to note that the data for the response variable, engineering % (overrun), does not follow a normal distribution and the researchers transformed it using a log-normal transformation to achieve desirable results. This allowed the researchers to use the variable in the model while passing the assumption of normality, a key assumption in the development of regression models as we discuss in Chapter III. The researchers found that variables pertaining to schedule had the most predictive ability of the 78 independent predictor variables analyzed in the study (Sipple et al., 2002).

From this study, we are able to take two main findings to further our own research. Most importantly, we learn what variables performed well to predict program cost and can explore using them in our own research. The most predictive variables were found to be as follows: maturity from MS-II (synonymous with MS-B), not using a major military contractor (i.e. not using Lockheed-Martin, Northrup Grumman, Boeing, Raytheon, Litton, or General Dynamics), and program acquisition unit cost (Sipple et al., 2002). This being said, there are a total of 78 predictive variables explained by this research which we can use to explore cost growth in our research.

Secondly, we find further evidence that schedule seems to be a major driver of cost in

DoD acquisition programs. This leads us to believe that the same variables used to predict schedule may be used to predict cost since they appear to be so intimately entwined. Additionally, this drives our belief moving forward into our analysis that there is a positive correlation between cost and schedule.

Relationship Between Cost and Schedule

The ability to combine two of the three sides of the acquisition program triangle, cost and schedule, into a comprehensive and usable tool for the acquisition community at large allows program managers to identify the programs that may have a higher risk of overrunning cost and schedule estimates. It is generally accepted knowledge that there exists a trade-space within the triangle that every program is confined to, yet no studies exist, to our knowledge, which attempt to quantify or analyze how any two of the different facets of the triangle interact outside of one NASA study. We fill this gap with our research using the cost and schedule analyses to determine how they interact. The following are the only prior research efforts to analyze cost and schedule in the same study.

In August 2014, at the annual NASA Cost Symposium, Burgess Consulting, Inc. gave a presentation titled Integrated Estimating Relationships in which they tie together three main aspects of NASA programs in order to predict the probability of a program accomplishing its cost or schedule goal. The research studies the interaction between the integrated relationships between cost estimating, schedule estimating, and mission phasing. Mission phasing is the outlay profile of the program, which corresponds to when funding is actually spent. A total of 37 NASA programs were used in the study and multiple regression models were developed to analyze the relationships between the three variables (Burgess & Krause, 2014)

Three estimating relationships were quantified by this study: Phasing estimating relationship (PER), Cost estimating relationship (CER), and Schedule estimating relationship (SER). The PER is the outlay profile of the program, quantifying how front-loaded or back-loaded the funding is. The CER is the total program cost from the system requirements review (SRR), a date that occurs prior to MS-B, through launch; while the SER is the time, in months, from SRR to launch. Each of these variables is computed using a multiple-regression equation to estimate the output based on four independent inputs. Unfortunately, since each input is specific to space acquisitions, and not found in all programs outside of space acquisition, we cannot analyze any of the variables in our own research. From this analysis, a set of tools was developed to give decision makers the ability to quantify trade-offs between cost, schedule, and phasing in their program. Additionally, this analysis allows the program manager to conduct a programmatic "health assessment" in which the estimating relationships are analyzed to determine if they fall within a standard deviation of the mean observed historical value (Burgess & Krause, 2014). Our research follows closely in line with their trade-off analysis.

The trade-off analysis conducted in this research relies on conditional analysis and normal probability curves. This research has developed a tool to predict the probability of a program meeting one of the estimating relationships given the other two. An example given in the presentation is "Given my project's budget profile & cost estimate, what is the probability it will be ready for launch by the need date?" (Burgess & Krause, 2014:14). From this research we can take a great deal of information to be used in our own. Conditional analysis is something that can be applied to our own research in order to add an additional layer of analysis for use by the decision maker. While this research is very similar to ours, there are many important differences. There are relatively few parallels that we can draw between space programs and the programs in our database, mainly regarding program constraints and unit quantity, meaning that our analysis is vastly different from theirs. Additionally, we only use cost and schedule, we do not use program phasing in this research.

The Institute for Defense Analysis (IDA) has conducted research for the DoD in various capacities. We identify two specific studies pertaining to cost and schedule trends from their pool of research. The primary study, conducted in 1989, specifically analyzed the effectiveness of acquisition initiatives that were present at the time. This research consisted of 82 programs, gathered from SARs, which met the main criteria of the research team of not being recent starts. The researchers used a regression analysis to determine which variables had the largest significant impact on cost growth. They found that development schedule growth, development schedule length, and production stretch (defined as procuring the same quantity over a longer period than planned) were the major drivers of total program cost growth (Tyson, Nelson, Om, & Palmer, 1989). This research stands as a prime example of program schedule driving costs.

The second study, conducted in 1994, expands upon the initial study from 1989 to identify cost and schedule growth patterns. According to their research, longer schedules mean higher costs, and cost growth made systems less affordable while simultaneously eroding congressional support. This study consisted of tactical missiles, tactical aircraft, and included 82 programs analyzed between Milestone II (currently referred to MS-B) and IOC (Tyson, Harmon, & Utech, 1994). Our main takeaway from this research is the relationship between schedule growth and cost growth. Tyson et al. (1994) developed two multiple regression models; one for tactical missiles and one for tactical aircraft, which use schedule and aspects of unit cost to predict total acquisition cost. While this research analyzes cost and schedule together, the schedule is not analyzed as a separate model, but, rather, is a variable to predict total cost growth. This suggests to us that there is a positive correlation between schedule growth and cost growth.

Summary

This review of relevant literature notes that there have been numerous studies performed before ours that help direct our efforts. What we have uncovered has shed light on the many methodologies that have been employed, as well as their findings under each aspect of our research. With the knowledge of prior research, we are able to confidently identify our starting point and strategy moving forward to our methodology.

One finding is clear after conducting this review; there is very little research linking together two of the most important aspects of a program into one analysis: the program's cost and schedule. By reviewing the literature, we now know that there is this need in the community and we can fill it. We gained the insight into how to structure our methodology, which is covered in detail in the ensuing chapter.

III. Methodology

Chapter Overview

In this chapter, we discuss the procedures used to conduct our research in seven separate sections. We begin with a discussion of our database—to include data sources, compilation, and assumptions. From there we move into defining our response variable for program cost. Next, we discuss the pursuit of predictor variables for cost and set the stage for how they are analyzed and selected. Then we discuss the multiple regression technique we use and experiment-wise error rate we accept. Afterwards, we outline the tests and procedures we must conduct to ensure our predictive model is stable and applicable to the data analyzed. From there we discuss the validation of our model and the reinsertion of the validation pool to create the final models. Lastly, we outline how we analyze the programs logistically to predict cost and schedule threshold overruns.

Database

To conduct our analysis, we first utilize the database built by Jimenez et al. (2016) in our study, as we build upon the research he conducted here at AFIT. We build upon the database employed in 2016 by using SAR data and filling in missing critical data points through data inquiries on the internet, specifically through the Air Force Magazine and Deagel.com, and using correspondence through the program offices at Wright-Patterson Air Force Base (WPAFB). In this section we identify what we use to identify and collect program data before outlining inclusion and exclusion requirements for our database. We continue by detailing how our complete database is populated; building off prior research as well as the additional research we conduct ourselves. We then touch on the use of a validation pool for our research before discussing several key assumptions we make regarding data collection and inclusion.

Selected Acquisition Report (SAR) Data

This research uses SAR data almost exclusively to populate the database we analyze and make inferences from. According to the DAU, the SAR is a comprehensive summary status report of a Major Defense Acquisition Program (MDAP) and is required for periodic submission to Congress; it includes key cost, schedule, and technical information in a standard format. Entry into the Defense Acquisition Management Information Retrieval (DAMIR) system is controlled, as it includes both classified and unclassified programs; therefore, for national security reasons, this research only includes unclassified information. The programs included within the database represent those of the highest interest to the government and public.

As noted by previous research discussed in the literature review, the data contained within the SARs is frequently used by the DoD to research cost and schedule growth in the acquisition environment. The database and reporting instructions for the SARs have been improved upon over the decades since its inception; however, prior research still notes difficulty in using the data. Most recently Jimenez et al. (2016) had to filter 80 programs from the database due to missing information regarding IOC alone—a very important program date. Research conducted into the shortfalls of the SAR database was conducted by Hugh (1992) to identify the issues and potential causes for those issues. The researchers identified the most notable problems to be exclusion of significant elements of cost and constantly changing guidelines, among others. Hugh (1992) concludes, however, that even though SAR data have a number of limitations, they are suitable for identifying broad based trends across program. With this in mind, the SAR database remains an appropriate source of data for the research we conduct.

Inclusion Criteria

The inclusion criteria for this research is any program in the DoD, to include all service branches, which have reported program information using program SARs. Additionally, this research focuses on only programs which are unclassified and reportable in the DAMIR database. These include high interest defense acquisition programs spanning from the 1960's to today.

Exclusion Criteria

This research has three exclusion criteria for the first portion of our analysis, as well as a fourth criteria that is applied only to the logistic analysis portion of this research. Each criteria is imperative to this research in order to create a robust and useful tool for use in the acquisition community. The criteria are outlined in detail in this section. Tables depicting the exclusion criteria, as well as the number of programs removed due to each criterion, are detailed in chapters four and five of this thesis.

The first exclusion criterion is that the program must contain a MS A date in the schedule portion of the SAR as well as contain corresponding cost information for MS-A. It is important to note that an exception to this is criteria is specifically detailed in the Key Data Assumptions portion of this chapter. The purpose of this exclusion criterion is to ensure that programs include information for three important predictor variables: MS-A to MS-B Duration, RDT&E $ (M) at MS-B Start, and % RDT&E Funding at MS-B Start. These variables are defined in Appendix B.

The second exclusion criteria is that the program SAR must contain a MS B date and corresponding funding information. This again pertains to the necessity of containing pre-MS

B data as a means to build a highly predictive model. Without the MS B date and funding information, we are unable to ascertain the duration of MS A or the funding spent up to MS B. Additionally, we are unable to calculate the projected funding needed to reach IOC, or the projected duration of MS B to IOC.

The third exclusion criterion is that all program SARs must contain an IOC date that has occurred prior to the latest reported SAR. We choose to include only these programs for several reasons. First, Jimenez's research uses IOC as a termination date for their predictive model and, as one objective of our research is to re-validate the model developed in his research, we continue to use this as the termination date. Second, the SAR database includes information on all programs of great importance to congress, even if the program was terminated. By requiring the programs we include to have an IOC date that occurs after the last SAR report date, and corresponding cost information, we assure that each program is completed at least up to this point. Third, Kozlak (2016) found that, on average, programs experienced 91% of their cost growth at IOC. While we do not know if this finding translates to schedule growth being similarly realized at this point, it leads us to believe IOC may be a good termination point for our schedule model. Additionally, by using IOC, our cost model has a predictive termination point.

The fourth exclusion criterion, which is only applied to the logistic analysis of this research, is that the program must have a SAR with a cost and schedule estimate within one year of MS B. This is critical to our logistic regression analysis because it gives us the current cost and schedule estimates, at MS B, to compare to the actuals at IOC. This, in turn, allows us to use the numbers to determine whether the program experienced a significant or critical overrun of their current MS B estimates and allows us to place them into "buckets" for analysis as explained later.

Database Population

Due to the dual nature of our research, we have two separate databases that are analyzed, one being a subset of the other. We outline, in detail, how each database is populated in Chapter IV and V of this thesis. The first database in Chapter IV includes 73 programs and was populated using past research, most recently Jimenez et al. (2016), as well as program SARs accessed through DAMIR. This database is built using three exclusion criteria. The second database in Chapter V includes 49 programs and is a subset of the first database. This database is built using four exclusion criteria.

Validation Pool

In order to test and validate that our multiple regression models perform the way they are intended, and to ensure we did not erroneously build a predictive model that only predicts the specific programs it was built on, we must randomly set aside 20% of the programs contained within our database to test the models after they are built. This means that 15 programs are not used to build the predictive models, but instead are used after the models are built to test how accurately the models perform when applied to new data. We discuss how this test is performed in detail in a later section. Once the model is validated against the validation pool and found to be sufficient, these programs are re-inserted into the whole database and the model is recreated using all programs to create the most representative model possible. It must be noted that the logistic regression models do not follow this validation method due to the small sample size.

Key Data Assumptions

There are several key assumptions made throughout this data gathering process. The most important assumption concerns the SAR not containing a Milestone A date outlined in the program schedule. In this case, if the funding profile contains a funding date at least one year prior to the start of Milestone B, we assume that Milestone A occurred in January of the earliest funded year. January is identified as the most appropriate month because continuing resolution authority (CRA) is a typical issue which the DoD has to contend with and, consequently, programs are unable to spend funding if they are a new start (which is the definition of a program at Milestone A) until a presidential budget is signed. This generally occurs in January, instead of October, and programs are allowed to start spending funding and sign contracts above what they were previously allocated. There are 15 programs in our database identified as not having a Milestone A date, but having pre-Milestone B funding.

To track these programs, and to detect if they are statistically different from the other programs in the database, we created a separate variable. This variable is called "no MS-A date" and its sole purpose it to determine if these programs can be included in the research due to this assumption.

Since many of the SARs within the database span different eras of DoD acquisition, several milestones and requirements follow different naming conventions than we use today. For our study, we consider Milestone II equal to Milestone B based on their corresponding definitions and only refer to this point in the program's life-cycle as MS-B in our study, which is consistent with former studies (Harmon, 2012). The same holds true for Milestone I and III being equivalent to Milestone A and C respectively.

We assume that, if the program created a prototype, it would be mentioned in that programs SAR. If a prototype is not mentioned, then it is assumed no prototype was created. This is because whether or not a program created a prototype is not a required reporting item and we have no other official data source to tell if one was ever created. Additionally, for aircraft programs only, if first flight is prior to MS-B then it is assumed this first flight was conducted by a prototype since research and development is not concluded at this point and that first asset must be a "proof of concept" asset.

It is assumed that the SARs are not only accurate and representative, but also all inclusive. This means that the programs identified in DAMIR as being applicable to this research are not only representative of DoD programs in general, but are accurate and unbiased by those who entered the information into the reports. It also means we assume all RDT&E and Procurement costs are included in the data.

Response Variables

A response variable is the dependent variable in an equation; in our case, it is the variable that we predict using several input, or predictor, variables. The costs within our database are expressed in Millions of dollars ($M) and would be analyzed the same in our model if not for the issue of homoscedasticity. As discussed in the ensuing chapter of this research, we must transform our response variable using the natural log in order to reduce the high level of variability in the cost of a program.

For our multiple regression model conducted in JMP®, our cost response variable is as follows:

- *Cost from MS-B to IOC BY17 ($M) [Regression Output]*
    - This variable states the actual cost (both RDT&E and Procurement) from the beginning of MS-B to IOC in BY17. This data is unavailable to the cost estimator at the time they are developing a cost estimate. This variable is transformed using a natural log and the output is re-transformed to give the

median output as expanded upon later in this thesis.

For our Logistic regression models we use three separate response variables, analyzed two ways each. The response variables will identify the programs that meet the requirements for being considered in the given response. These variables are as follows:

- *Group I [Regression Output]*
    - This variable states that the program cost and schedule estimate at MS B is within the threshold limit of the actual cost and schedule duration at IOC. There are two thresholds analyzed, one being significant (15% over the current APB) and the other is critical (25% over the current APB). This is used to label these programs as not oing their current Acquisition Program Baseline (APB) threshold when compared to the actual at IOC

- *Group II [Regression Output]*
    - This variable states that either the program cost or schedule estimate, but not both, at MS B overruns the threshold limit of the actual cost or schedule duration at IOC. There are two thresholds analyzed, one being significant (15%over the current APB) and the other critical (25% over the current APB). This is used to label the programs as overrunning only one, non-identified, threshold when comparing their current APB versus the actual at IOC.

- *Group III [Regression Output]*
    - This variable states that bot the program cost and schedule estimate at MS B overrun the threshold limit of the actual cost and schedule duration at IOC. There are two thresholds analyzed, one being significant (15% over the current APB) and the other critical (25% over the current APB). This is used to label the programs as overrunning both thresholds when comparing the current APB to the actual at IOC.

Predictor Variables

Leveraging the research we analyzed in our literature review we identify many candidate predictor variables for predicting program cost which can then be used in the logistic analysis for cost and schedule overruns. We recognize that the same variable may be found to be predictive of both cost and schedule; as the models are calculated independently, this is of no concern. It is imperative that the predictor variables contain a logical relationship to the variable which they predict in order for the model to be of any use to the community. By following only logical leads, we maximize our chances of discovering causal relationships instead of stumbling upon simple corollary predictors that are indefensible and a matter of simple chance.

In order to protect against non-constant variance in the cost model, due to the large range in dollar values present in the predictor variable, we analyze the histogram of program costs in order to determine natural break points in the data. This allows us to discretize the data by placing the programs into "bins" labelled as small, medium, large, and extra-large programs depending on the break points determined in the data. In order to remove any researcher bias from the analysis of the histogram, the interquartile values and the mean are used to determine the values that are contained within each bin.

We clearly define each one of our predictor variables examined in this study, to include the units, variable type, and, if necessary, how this variable is calculated or obtained. To be included in our final regression model, described in detail in Chapters IV and V of this thesis, the predictor variables must also demonstrate a level of significance using the Holm-Bonferroni Correction technique at $\alpha = .10$ in addition to being logically related to the response variable. Each predictor variable is found across all programs in our database. The predictor variables, with definitions, are outlined in appendix B.

Multiple Regression Analysis

The culmination of the entire data gathering process is building the regression models to predict the response variable. Analysis is performed individually for each model, and the results are independent from one another. To begin this process we test the desired response variable and include every predictor variable we have gathered following a mixed direction stepwise regression analysis. We set our Type I error to be 0.1 in lieu of 0.05 as this research is more exploratory instead of explanatory. This technique allows us to identify the most predictive variables that we have at our disposal immediately.

The variables are scrutinized by their individual p-values when determining which variables to keep in the final model. We analyze them using the Holm-Bonferroni criteria, which is a technique to ensure we do not violate the overall experiment error rate for the model; again, our Type I error is $\alpha = .10$. This method allows us to avoid over-fitting our regression output by including more predictor variables than are relevant without excluding others that are borderline variables.

When analyzing the performance of a regression output, two particularly good measures of how well the model performs are the R^2 and Adjusted R^2 values. The R^2, also known as the coefficient of determination, is a measure of how well the model predicts the regression output. In other words, it is a measure for how much variability present in the data is explained by the regression equation. The Adjusted R^2 value is a modified version of R^2, the difference being that Adjusted $\boldsymbol{R^2}$ accounts and adjusts for the number of predictor variables in the model. These two outputs are what the performance of our models are judged by. A value of one means that the model perfectly predicts the output and zero variability is left unexplained, while a value of zero, conversely, means that there is no predictive capability in the model and 100% unexplained variability present. Naturally, the aim of this research is to obtain R^2 and Adjusted R^2 values as close as possible to one.

Once the variables are finalized, we build the model, use the outputs to perform the requisite tests identified, and subsequently validate the model as described later in this chapter. The model is only valid for the purposes of this research if it passes all identified tests and is validated by the validation pool using the identified techniques. We follow this process for both the schedule and the cost model.

Checks and Tests Performed

In order to build relevant and statistically significant multiple regression models, several statistical tests must be performed and techniques must be applied. By performing the following tests, which we overview in this section and apply in Chapter IV, we ensure that our model is robust. These tests are performed after a suitable regression model is created; their purpose is to validate that the model is statistically sound and that it is applicable to the data used to create the model by isolating and examining any peculiarities in the data that may invalidate our model.

Variance Inflation Factors

Once the model is built, the analysis commences with analyzing the variance inflation factor (VIF) scores of each predictor variable that is deemed statistically significant. The VIF measures how much multicollinearity has increased the variance of an estimate, meaning that it measures and indicates the magnitude that a single predictor variable influences the outcome of a multiple-regression model (Stine, 1995).

Cook's Distance

This test is used to detect overly influential data points within the dataset that are possibly skewing the results. Cook's Distance is commonly used in multiple regression analysis to interpret each data point's influence on the regression results and can easily highlight outliers through a graphical interface and a corresponding "score" for each data point between 0 and 1. An overly-influential data point can be potentially harmful in a regression model, due to the fact that it over-fits the regression output to include that one instance. In essence, it influences the model to behave more like the one observed outcome than the population as a whole. When utilizing Cook's Distance, we are alerted to any possible points which need to be analyzed more closely by a score of between .1 and .5. It is probable cause for removal of a data point if the Cook's Distance score is greater than .5—meaning that the model is entirely dependent on that one point and may not be applicable to any future scenarios (Cook, 1977). This would clearly render our model useless to the acquisition community.

Studentized Residuals

The histogram of the studentized residuals is analyzed to detect any potential outliers in the data. For our research, any data is considered to be a potential outlier if the studentized residual is greater than three standard deviations from the mean. This prompts us to analyze the points further to determine if there is an issue with the data, such as a transcription error or a program, which is not properly accounted for in our data set, and can indicate that a data point may be removed. Any points that reside three or more standard deviations from the mean are analyzed, and their inclusion or removal is discussed and defended in detail in Chapter IV.

Shapiro-Wilk's Test

In order for our multiple regression models to be valid and, therefore, useful to the community they are designed to serve, they must pass the Shapiro-Wilk's (S-W) test of the assumption of normality. The S-W tests whether the residuals in our set come from a normal distribution or not. The assumption that the residuals come from a normally distributed population is one of two key assumptions to building a regression model and, if our models do not pass this test, they are considered unusable to predict the outcomes. The null hypothesis for the S-W test is that the residuals from our model are normally distributed; the alternative hypothesis is that they are not. We test this at a threshold of $\alpha = .05$. If the p-value for the test is larger than .05, then we satisfy the assumption of normality for our models (Neter et al., 1996:111).

Breusch-Pagan Test

The second of the two key assumptions of any multiple regression model is that it contains constant variance. The Breusch-Pagan (B-P) test is used to statistically prove whether a model exhibits constant variance or not. In order for our models to be valid in predicting their respective outcomes, the variance from the errors in the model must not be dependent on the independent variables (Neter et al., 1996:239). This test is used to determine whether heteroscedasticity is present in the model, which identifies the variance in the model created as being non-constant and, therefore, identifying an issue with using those predictor variables in our model.

In order to pass the assumption of constant variance we use a p-value of $\alpha = .05$, meaning we must obtain an output from the B-P greater 0.05. This ensures our model is robust and the variables we use are truly predictive of the output.

Validation of Models

The closing test for our model is to test it against the validation pool, which contains 20% of the total programs analyzed, that was removed prior to beginning the multiple regression analysis. To conduct the validation we must measure how well our model predicts the individual outcomes of each program in the validation pool against the actual observed outcomes. To do this we run the validation pool through the cost and schedule models to collect the outputs. We analyze the Absolute Percent Error (APE) of each program's outputs. The APE is equivalent to: $\frac{Observed\ Value - Predicted\ Value}{Observed\ value}$. From the APE values we calculate the Mean APE (MAPE) and Median APE (MdAPE) values. We compare these values to the

MAPE and MdAPE of the programs that the models were built on. They should be similar if the models are predicting properly.

Once the values are compared an actuals by predicted bivariate plot is created for both the validation pool as well as the model building pool. These plots are analyzed for behavioral consistency between the two data pools. If they pass scrutiny then the predictive models are complete; the data pools are combined and the final models are created.

Combination of Cost and Schedule

The culmination of this research is to analyze the cost and schedule of a program from a longitudinal standpoint. This means that we identify each program for meeting their cost and schedule estimates, from MS B to IOC, compared to the actual cost incurred from MS B to IOC using the threshold values identified in a Nunn-McCurdy breach. These breach thresholds are 15 and 25% when considering the estimate at MS B to be the current APB.

We analyze these programs and their outputs based on four categories. Each threshold is analyzed independent of the other and, while one is a subset of the other, we do not analyze them simultaneously. Programs enter into each category based whether their MS B estimates overrun the actual multiplied by the threshold. The first group includes programs that have not overrun either schedule or cost. The second group includes only programs that have overrun either cost or schedule, but not both. The final group includes programs that have overrun both cost and schedule. What this allows us to do is analyze and identify common factors inherent in each program that may lead them to be included in either group. The application of this knowledge allows decision makers to identify potential problem programs as well as potential over performers. There is also the potential to identify characteristics that may lead to only a cost or only a schedule overrun, but not both.

To conduct our logistic regression analysis, we not only have to identify which bins each program belongs to, we must transform each continuous variable into a nominal variable, coding the positive response with a one and the negative response with a zero. This simplifies the analysis and application of the logistic models, allowing us to readily interpret the odds ratios. To do this, we analyze each continuous variable for the programs contained in each bin separately to identify points that indicate a trend in the data. This process is explained in further detail in Chapter V of this thesis.

Summary

Using the findings in our literature review, we built a set of predictor variables that form the backbone of our analysis. This enables us to intelligently defend the use of variables we find predictive of either response variable and argue a logical link between them. We outlined our collection of data in order to develop a more robust dataset and ensure applicability to a wider user community. We also provide systematic instructions for the data analysis and model-building process, which enables the process to be reconstructed while also defending our procedures. In Chapter IV, we put the theory into action to demonstrate the results of our cost model analysis. In Chapter V we use the estimates and thresholds to categorize the programs and use logistic regression to calculate the probability of a program identifying with each group. In Chapter VI we discuss our results and how they answer our research questions, as well as what our recommendations are for using this research and any future research, related to this research, that should be accomplished.

Modeling Median Will-Cost Estimates for Milestone B to IOC in Defense Acquisition Programs

Abstract

The introduction of "should cost" in 2011 requires all Major Defense Acquisition Programs to create efficiencies and improvements to reduce a program's "will-cost" estimate. Realistic "will-cost" estimates are a necessary condition for the "should cost" analysis to be effectively implemented. Due to the inherent difficulties in establishing a program's will-cost estimate, we propose a new method to infuse realism into this estimate. Using historical data from 73 Department of Defense programs as recorded in the Selected Acquisition Reports, we utilize mixed stepwise regression to predict a program's cost from Milestone B (MS B) to initial operational capability (IOC). Our presented model explains 83% of the variation in program acquisition cost. Significant predictor variables include: projected duration (months from MS B to IOC); the amount of Research Development Test and Evaluation (RDT&E) funding spent at the start of MS B; whether the program is considered a fixed wing aircraft; whether a program is considered an electronic system program; whether a program is considered ACAT I at MS B; and program size relative to the total program's projected acquisition costs at MS B.

Modeling Median Will-Cost Estimates for Milestone B to IOC in Defense Acquisition Programs

Introduction

On June 15, 2011 the Under Secretary of Defense for Acquisition, Technology, & Logistics (USD (AT&L)) directed the Military Departments and Directors of Defense Agencies via a memorandum (see Appendix A) to implement Will-Cost and Should-Cost management for all Acquisition Category (ACAT) I, II, and III programs. In this memorandum, the USD (AT&L) reiterates that the Departments will continue to set program budget baselines using non-advocate Will-Cost estimates. A Will-Cost estimate uses traditional cost estimating techniques (e.g. analogy, bottom-up, parametric, etc.) to estimate the most-likely cost of a program in order to establish a reasonable budget baseline and acquisition program thresholds. However, the USD (AT&L) also "challenges program managers to drive productivity improvements into their programs during…program execution by conducting Should-Cost Analysis," which involves, "identifying and eliminating process inefficiencies and embracing cost savings opportunities." (Carter & Mueller, 2011, p. 16). The Should-Cost estimate therefore deviates below the Will-Cost estimate to develop a realistic price objective for negotiation purposes and subsequent savings against the Will-Cost estimate.

Additionally, the USD (AT&L) states in the same memorandum that, "the main problem with the will-cost estimate isn't in the numbers or how it was reached; the problem is that once the will-cost estimate is derived and the budget for the program is set, historically, this figure becomes the "floor" from which costs escalate, rather than a "ceiling" below which costs are contained—in many ways creating a self-fulfilling prophecy of budgetary excess.” (Carter & Mueller, 2011, p. 16). We suggest that perhaps there is a better way to infuse realism into a Will-Cost estimate such that it becomes a middle of the road estimate from which to work from in the Should-Cost approach rather than the floor.

Therein lies the crux of the problem; how does one go about generating a median ‘Will-Cost’ estimate? Defense acquisition programs expand the frontiers of today’s technology to develop new and innovative systems that provide an asymmetric advantage on the battlefield. As a result, there are inherent uncertainties and risks associated with Department of Defense (DoD) acquisitions. These realities are manifested in the derivation of the program’s cost estimate. To combat risk and uncertainty, cost analysts tend to err on the side of caution and unnecessarily inflate the cost estimate to capture more of the risk. This method needlessly ties up precious resources that may be better placed elsewhere. By contrast, building an overly aggressive cost estimate may free up resources to be placed elsewhere. However if this estimate is exceeded, decision makers could take critical funding from other programs or force a program manager to delay the program until additional funding can be secured.

In order to combat these issues, programs should strive for a realistic, middle ground point—essentially an empirically validated baseline. The use of historical data allows the acquisition community to unbiasedly analyze and to estimate what a program would cost in relation to other similarly completed programs. This estimate then becomes a powerful tool from which the user can identify a target cost for a given program. This estimate also serves as a benchmark to identify whether a cost estimate is reasonable given what has occurred in the past. From this estimate, mitigation of risks associated with over- and under-estimating program costs may be achieved resulting in a more efficient allocation of resources. Thus, we propose a new empirically-based model for determining Will-Cost estimates in DoD acquisition programs.

Past Research and Database Creation

Our research is intended to build an empirically derived model to predict median Will-Cost estimates for DoD acquisition programs. We utilize prior research to identify potential explanatory variables in our model and establish the basis for creating our dataset. Our literature review spans the change in acquisition taxonomy from Milestone I, II, III to Milestone A, B, C. For this study, we consider Milestone I, II, and III to be equivalent to Milestone A, B, and C, respectively. This is consistent with prior literature findings that the naming convention has simply altered over time without tangible changes in definition or substance (Harmon, 2012; Jimenez, White, Brown, Ritschel, Lucas, & Seibel, 2016). Prior to relaying our data collection process, we first discuss recent studies pertinent to our research. Using these studies, we build the foundation for how we conduct our research into the relative program characteristics that predict program cost.

Jimenez et al. (2016) developed a schedule duration prediction model for defense acquisition programs using pre-MS B data; we leverage their research to identify explanatory variables for investigation and an initial dataset from which to draw upon. Their analysis concluded the following variables were significant in establishing an empirical benchmark for "Should Schedule" estimates: amount of RDT&E dollars (in millions) at MS B start, the percent of RDT&E funding at MS B start, whether a program is a modification, and whether a program has a MS B start in 1985 or later to be significant variables. Although they explored and adopted significant variables to predict program schedule using pre-MS B data, they also considered a plethora of explanatory variables that were ultimately deemed statistically insignificant; we also consider these variables.

Brown, White, Ritschel, & Seibel (2015) first identified the MS B start in 1985 or later as an explanatory variable. They demonstrate that programs with a MS B start date in 1985 or later have a statistically significant change in their expenditure profile. These programs tend to expend a greater percentage of their obligations by the program's mid-point than the programs that start prior to 1985. Although not conclusive, Brown et al. (2015) hypothesized that the reason for this significant shift is due to the President's Blue Ribbon Commission on Defense (often referred to as the Packard Commission) and the acquisition reforms that occurred due to the recommendations of the commission.

Similar to Jimenez et al. (2016), Deitz, Eveleigh, Holzer, & Sarkani (2013) analyzed acitivities prior to MS B. They examined the importance of developing a robust Analysis of Alternatives prior to MS B and the effects that an analysis of this nature may have on program success. Their findings suggest that while only 10% of a program's lifecycle cost was invested prior to MS B, 70% of a program's lifecycle costs are committed by this milestone (Deitz et al., 2013). This suggests to us that pre-MS B data may be very important to predicting program cost. However, this also limits data collection because pre-MS B reporting is not mandatory for all acquisition programs and therefore the cost and schedule data is unavailable in some instances. Jimenez et al. (2016) also experienced such a limitation.

Looking slightly further back in the literature, we find other pertinent studies that present possible explanatory variables to consider. Foreman (2007) researched methods to improve cost and schedule growth estimates by including longitudinal variables that account for changes that take place over time. His research built upon the database initially comprised by Sipple, White,

& Greiner (2004) and subsequently modified by Lucas (2004) and Genest & White (2005). Sipple et al. (2004) found the most important predictive variables of cost growth to be were MS C to IOC duration and an indicator variable for a MS C slip.

The aforementioned researchers have identified numerous variables for investigation on whether they will be predictive of program cost. The complete list is in Appendix B. This list also gives us our data inclusion and exclusion criteria. The initial data inclusion criteria is any program in the DoD (i.e., all service branches) which have reported program data using the Selected Acquisition Reports (SAR). Additionally, they must be unclassified and reported within the Major Defense Acquisition Program (MDAP) and pre-Major Defense Acquisition Program (pre-MDAP) section of the Defense Acquisition Management Information Retrieval (DAMIR) database.

For a program to be considered in our study, it must satisfy three criteria; otherwise, it is excluded for this research. The first requirement is that the program SAR must contain a MS A date or funding at least one year prior to MS B – we interpret the pre-MS B funding as indicating the year in which MS A may have occurred. This requirement is due to the pre-MS B data being found predictive in the literature review. Unfortunately, this requirement also results in a great deal of programs being ineligible for inclusion because a lack of reporting requirements prior to MS B. This is not unexpected considering a program is not official until meeting MS B.

We are able to include 15 programs in our dataset by making the following assumption when there is no MS A date provided: if there is funding in the funding profile at least one year prior to MS B, then MS A occurred in January of the year in which funding was first received. We test this assumption to ensure they are not statistically different from the others prior to inclusion in the final dataset; as we’ll see in the methodology section, they are deemed statistically equivalent.

The second exclusion criteria is that the program SAR must contain a MS B date and corresponding funding information. This again pertains to the necessity of containing pre-MS B data as a means to build a highly predictive model. Without the MS B date and funding information, we are unable to ascertain the duration of MS A or the funding spent up to MS B. Additionally, we are unable to calculate the projected funding needed to reach IOC or the projected duration of MS B to IOC.

The third exclusion criteria is that the program SAR must contain an IOC date that occurred prior to the last reported SAR which indicates that the program is complete up to IOC. This is important to our research as it gives us a termination point to estimate and ensures we are not using projected values as actuals in our model. IOC is a very important date in a program as it signifies the point in time when the user community can first begin to realize the benefits from the investment in the program.

As previously discussed, our dataset starts with the 56 programs in the database built by Jimenez et al. (2016). We augment this database by analyzing defense program SARs from the DAMIR system.. The program SARs contain program funding, schedule, and performance information relative to our research. Using our stated inclusion criteria, we add 187 programs to the initial 56. Then using the exclusion criteria, we remove 170 programs for a net change of 17; this results in a final program count of 73. Table 1 demonstrates inclusion and exclusion criteria used in this research. The list of all included programs is located in Appendix C.

Table 1: Program Inclusion / Exclusion Criteria and Counts

Inclusion/Exclusion Criteria	Programs Added	Program Removed	Program Count
Jimenez's Starting Database	56		56
DAMIR Query and Addition	187		243
Double Count Adjustment		29	214
IOC Occurs after Last SAR		61	153
Missing Milestones A or B		74	79
Missing IOC		4	75
Classified		2	73
Total Remaining			73

The data that we use for our analysis includes both actuals and projected values from the SARs. We use the latest available program's SAR to record the actual cost from MS B to IOC as the response variable in the model. In order to develop a useful predictive tool for the acquisition community, we must only use projected cost and schedule data at MS B since this is the only data the user of our regression model will have at their disposal at that time. This limitation ensures that we are not overly influencing our model by using data which will be unavailable to the user when they predict the actual program cost.

In order to implement this limitation, we retrieve projected cost and schedule data from the SAR corresponding to the year in which MS B occurred or, if that SAR is unavailable, the earliest available SAR. This allows us to use projected values to predict a program's cost from MS B to IOC, the same as if we were in a program office attempting to estimate the cost of our program independent of this research.

Methodology

To arrive at the presented model (explained in the next section), we use a mixed direction stepwise approach to screen for the most predictive variables, and then finalize the model using ordinary least squares (OLS). For our regression model, the response variable is the natural log of the acquisition cost (defined as the RDT&E and Procurement costs) from MS B to IOC. We transform the response variable using a natural log function in order to mitigate against heteroscedasticity due to the large range of actual costs—without transforming the OLS residuals, we would have failed the assumption of constant variance at a level of significance of 0.05. To ascertain the actual cost estimate from the OLS model we re-transform the predicted output back to actual cost (in millions of BY 17 dollars) by calculating $e^{OLS\ Output}$. This transformed model results in a median estimate of Will-Cost since this back-transformation equates to the median in the original space (Carroll & Rupert, 1981; Tisdel, 2006).

As noted earlier, Appendix B lists and defines the predictor variables analyzed—recall that the predictor variables must be available to the cost estimator and relevant to program cost in order to be considered in this research. As such, only actual data is used up to MS B, and projections are used after MS B. In order to be included as an explanatory variable, it must have occurred in at least one of the 73 programs in our database. Additionally, to eliminate the effects of inflation we convert all funding variables to BY17 using the 2016 OSD inflation indices.

We use JMP® Pro 12 for our statistical analyses and adopt an initial overall experiment wise Type I error of 0.1 due to the exploratory nature of this study. To be consistent with this level of significance, we use a p-value threshold of 0.1 to enter and exit the mixed direction stepwise regression model. Once the initial variables are identified by the stepwise procedure, we then use OLS to finalize the regression model. We now lower the overall Type I to be 0.05 and require each predictor variable to be statistically significant according to the Holm–Bonferroni method, which counteracts the problem of multiple comparisons (Holm, 1979).

Prior to conducting the variable selection procedure, we randomly select 20%, or 15, of the 73 programs and set these aside for utilization as a validation set. We use the remaining 58 programs for the stepwise and OLS regression analysis. After validating our selected model, we perform another mixed stepwise analysis using the entire dataset of 73 programs to determine if we inadvertently left out a predictive variable.

For our model to be considered viable, we must verify the standard OLS assumptions. First, we use only one final SAR per program to record the actual cost and schedule information in addition to an earlier SAR for that program to ascertain the projected information—this assures independence. To assess the assumptions of homoscedasticity and normality of model residuals, we conduct a Breusch-Pagan (B-P) and Shapiro-Wilk (S-W) test, respectively, at a level of significance of 0.05. To assess multicollinearity between the predictor variables we examine their variance inflation factors (VIF). The VIF score must be below 10 in order for us to say with confidence that multicollinearity is not a factor in our model. To guard against a program being overly influential to the outcome, we employ Cook's Distance. A Cook's D value greater than 0.5 indicates the model is being overly influenced by a variable in one of the programs (Neter, Kutner, & Nachtsheim, 1996, p. 381).

After all the underlying model assumptions are assessed and passed, we test our resultant model against the validation pool using descriptive and inferential measures. Regarding descriptive measures, we compute the absolute percent error (APE), which is the absolute difference between the true cost between MS B and IOC and the predicted cost divided by the true cost for each program. Note, the true and predicted costs are evaluated in the natural log space. Using these APE values, we then calculate the median and mean APEs (MdAPE and MAPE, respectively). We calculate these for both the validation and modeling programs and compare the values. Finally, we investigate whether the untransformed predicted values truly reflect the median value or a baseline estimate for Will-Cost by investigating how the true program cost compare to the predicted program cost.

Analysis

Using mixed stepwise regression on the modeling set of 58 programs, we develop a preliminary model—Table 2 highlights this model. The presented model has an R^2, which represents the amount of variability in the data explained by the model, of 0.82. We calculate the APE values for this model which results in an MdAPE and MAPE of 0.050 (5.0%) and 0.059 (5.9%), respectively, for the model building set. For the validation set, we obtain an MdAPE and MAPE of 0.056 (5.6%) and 0.079 (7.9%), respectively. Although the validation set is slightly higher than the model building set, all of the absolute percent errors are less than 10% suggesting the model is performing well.

Table 2: Preliminary Ordinary Least Squares Model

Predictor Variable	Estimate	P-Value	Standardized Estimate	Variance Inflation Factor
Intercept	5.731	< 0.0001	N/A	N/A
Projected MS B to IOC (months)	0.0114	0.0033	0.199	1.13
RDT&E $ at MS B Start	0.00029	0.0003	0.297	1.56
Fixed Wing	0.620	0.0037	0.199	1.17
Electronic System Program	–0.732	0.0142	–0.216	1.96
ACAT I	0.837	0.0346	0.160	1.47
Large Program	0.747	0.0018	0.251	1.58
Extra Large Program	1.205	< 0.0001	0.397	2.16

With respect to the inferential measures, Table 2 reveals that all VIF scores are below or close to two indicating little to no evidence for multicollinearity. The preliminary model also contains no Cook's D score above 0.50 (highest value is approximately 0.10). This suggests no overly influential data points affecting the p-values of our explanatory variables. Model residuals pass both assumptions of normality and homoscedasticity with p-values of 0.25 and 0.92 for the S-W and B-P tests, respectively. Lastly, all explanatory variables are individually significant at the comparison-wise error rate under the Holm–Bonferroni criteria (Holm, 1979).

With the model being deemed internally valid, we combine all the data together to update model parameter values using OLS and lowering the overall Type I error rate to 0.05. Table 3 shows the updated model. Stepwise failed to detect any additional predictor variables (at the overall familywise error rate of 0.05 level of significance) and the resultant model described in the next section is our final model. The resultant model has an R^2 of 0.83 with an MdAPE and MAPE of 0.057 (5.7%) and 0.062 (6.2%), respectively. This means that the presented model has a relative error of between 5.7% and 6.2% of predicting the natural log of the program cost from MS B to IOC. After back-transforming to the original values of program cost from MS B to IOC, approximately 50.7% of the 73 programs in our database had a true program cost exceeding the predicted cost while 49.3% had less. Theoretically, this ratio should be 50% / 50%. The empirical percentages suggest our presented model is performing as expected.

To prevent model extrapolation, the ranges in which this model is useful for the two continuous variables must be consistent with the bounds of the programs used within our analysis. For projected duration from MS B to IOC the lower bound is 28 months while the upper bound is 129 months. For RDT&E funding ($M) at MS B Start (BY17), the lower bound is $4.43M while the upper bound is $5979.4M. Using this model outside of these ranges will invalidate the results. We now discuss the statistically significant predictor variables. All of these variables are available to the cost estimator at the time the estimate is calculated (which is intended to be post-MS B).

Table 3: Final Ordinary Least Squares Model

Predictor Variable	Estimate	P-Value	Standardized Estimate	Variance Inflation Factor
Intercept	5.449	< 0.0001	N/A	N/A
Projected MS B to IOC (months)	0.0108	0.0021	0.170	1.09
RDT&E $ at MS B Start	0.00026	0.0007	0.220	1.50
Fixed Wing	0.561	0.0039	0.165	1.19
Electronic System Program	–0.635	0.0061	–0.191	1.77
ACAT I	1.151	< 0.0001	0.251	1.38
Large Program	0.758	0.0004	0.232	1.51
Extra Large Program	1.461	< 0.0001	0.439	2.12

- (Projected) MS B to IOC Duration—Continuous Variable

  The parameter estimate of this variable is 0.0108 which is multiplied by the number of months the program estimates to spend from MS B to IOC. This duration does not necessarily correlate to the level of technology or technological maturity being employed, but, rather, indicates the cost of time in DoD acquisition.

- RDT&E Funding ($M) at MS-B Start (BY17)—Continuous Variable

  The parameter estimate associated with this variable is 0.00026, which is multiplied by the actual, non-transformed RDT&E funding spent prior to program entrance into MS B. Since the amount of funding spent at this point is additive to total program cost, we suggest that the amount of funding spent prior to MS B is indicative of the projected size and scope of the entire program. This variable could indicate a greater investment in newer technology prior to MS B, which typically results in higher costs over the entire program life due to integrating and further maturing this technology.

- Fixed Wing—Binary Variable

  The parameter estimate associated with this variable is 0.561 and will be multiplied by one for every aircraft (excluding helicopters) program estimate conducted. The positive parameter estimate indicates that aircraft programs sans helicopters appear to be more expensive in general in contrast to other DoD platform programs. We hypothesize this effect as an artifact of complexity associated with stealth, avionic, and engine capabilities of today's modern aircraft, regardless of branch of service.

- Electronic System Program—Binary Variable

  The parameter estimate associated with this variable is –0.635 and will be multiplied by one for any program that is considered an electronic system program. The negative parameter estimate indicates these programs are statistically significantly cheaper to acquire than the other program types. Bolten, Leonard, Arena, Younossi, & Sollinger (2008) also concluded that electronic systems are historically cheaper.

- ACAT I—Binary Variable

  The parameter estimate for this variable is 1.151 and is multiplied by a value of one for any program considered to meet ACAT I funding estimate requirements at the start of MS B. This variable being additive to program cost is logical due to the nature of ACAT

I programs and the dollar costs associated with these DoD acquisitions.

- Large Program—Binary Variable

  The parameter estimate for this variable is 0.758 and is multiplied by a one if the program being estimated projects to have a total program acquisition cost greater than $7B (BY17) (RDT&E and Procurement) from MS A to program conclusion but less than or equal to $17.5B (BY17). This value is estimated at MS B and was calculated using the 50% interquartile from a histogram analyzing total projected program acquisition cost. The additive nature of this variable adjusts for large DoD acquisition programs.

- Extra Large Program—Binary Variable

  The parameter estimate for this variable is 1.461 and is multiplied by a value of one if the program acquisition cost from MS A to IOC is projected to be greater than $17.5B (BY17). This value is estimated at MS B and was calculated using the 75% interquartile from a histogram analyzing total projected program acquisition cost. The additive nature of this variable adjusts for the largest DoD acquisition programs, such as the F-35 and F-22.

Discussion and Conclusion

Table 4 presents the relative percentage contribution of each variable included in the final model. The smallest relative contribution is 9.9% for Fixed Wing aircraft while the largest relative contribution is 26.3% for Extra Large programs. Besides these variables, there is low variation between the remaining predictor variables in the presented model. This suggests that the explanatory variables are relatively similar with respect to affecting the true program RDT&E and Procurement costs.

Table 4: Predictor Variables and Their Relative Contribution to the Model

Explanatory Variable	Relative Contribution
Projected MS B to IOC (months)	10.2%
RDT&E $ at MS B Start	13.2%
Fixed Wing	9.9%
Electronic System Program	11.5%
ACAT I	15.0%
Large Program	13.9%
Extra Large Program	26.3%

In addition to the significant predictor variables, it is also important to address the non-significant variables that were expected to contribute to our model analysis. The MS C Slip and Duration From MS C to IOC variables that were found to be significant in predicting cost growth by Foreman (2007) were not significant in predicting program cost from MS B to IOC. This finding allowed us to include the three satellite programs in our database since they do not have a MS C date within their SAR reports. Additionally, as was found by Jimenez et al. (2016), we do not find any statistical significance in branch of service, and the only program type that is statistically different is Electronic Systems. Lastly, modification, prototype, concurrency planned, and MS B start date in 1985 or later were not found to be statistically significant variables.

As with any statistical model, there are limitations to our regression model. Principally, this model is based on data collected from SAR reports that sometimes contain incomplete information. Ultimately, the model is only as good as the data itself. The availability of pre-MS B data was a large constraint on the data building process and limited which programs could be included. Additionally, the search parameters used in DAMIR may have inadvertently removed useful programs from our study which might have influenced any number of other variables to be significant.

One significant limitation is the high level of variability in the definition of IOC. Our model uses IOC as a termination point due to the importance of this milestone in a program as well as the availability of the date. In the programs we analyze, the number of units considered for attaining IOC varies greatly. Achieving IOC is determined individually for each unique program based on an initial cadre of operators, maintainers, and support equipment that can employ and sustain the system in an operational environment. For programs like satellites, submarines, and ships, IOC is generally considered to be very few, or even one single, unit. In the case of missile programs, IOC could be in the hundreds. This drives a level of known variability within our model that could be better accounted for by using a more structured and universal definition for IOC; this could be a topic for future research.

Accurately predicting program cost is both an art and a science. Achieving accurate estimates during the early stages of a program's lifecycle is an unenviable task, and one can be certain that the estimate will be wrong. However, deriving an estimate that is close to the final actual cost is crucial to improving the allocation of scarce resources. What our model provides is the empirical portion of the estimating process to ascertain the Will-Cost for a program. We provide this tool to the DoD acquisition community primarily as a method to check the assumptions and realism of their program office estimate. Being able to build a program cost estimate and turn to our statistically built and tested model for validation will be invaluable for the community because it will allow for an injection of increased realism into the cost estimating process. Realism in the Will-Cost median estimate is crucial to the success of Should-Cost analysis.

Drawing a difference between our research and prior research, the most notable difference is the model output. Our research and model focuses on building an empirically based estimate for program cost between MS B and IOC in order to serve as a realistic benchmark (the median value) for what programs Will-Cost. Program managers can then adopt "should cost" efficiencies to reduce cost further. We believe that modeling an output that will serve as an actual point estimate is valuable as a crosscheck tool for the user community. It gives the user a benchmark based on historical data against which the program measures its progress. The model also supports the "will-cost and should-cost" requirement levied in 2011 by providing an objective and defensible cost for what a program should actually cost based on what has been achieved in the past. Ultimately, a quality will cost estimate provides a starting point for program managers to examine processes and find efficiencies that lead to reduced program costs.

References

Birchler, D., Christle, G., & Groo, E. (2011). Cost implications of design/build concurrency. *Defense Acquisition Research Journal*, 18(3), 237–246.

Bolten, J.G., Leonard, R.S., Arena, M.V., Younossi, O. & Sollinger, J.M. (2008). *Sources of Weapon System Cost Growth: Analysis of 35 Major Defense Acquisition Programs (MG 670)*. Santa Monica, CA: RAND Corporation.

Brown, G. E., White, E. D., Ritschel, J. D., & Seibel, M. J. (2015). Time Phasing Aircraft R&D Using the Weibull and Beta Distributions. *Journal of Cost Analysis and Parametrics*, 8(3), 150–164. http://doi.org/10.1080/1941658X.2015.1096219

Carroll, R. J., & D. Rupert (1981). On prediction and the power transformation family. *Biometrika*, 68(3), 609-615. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.536.2459&rep=rep1&type=pdf

Carter, A. B., & Mueller, J. (2011). Should Cost Management: Why? How? *Defense AT&L*, 40(5), 14–18. http://www.dau.mil/pubscats/ATL%20Docs/Sep-Oct11/Carter_Mueller.pdf

Deitz, D., Eveleigh, T. J., Holzer, T. H., & Sarkani, S. (2013). Improving program success through systems engineering tools in the pre-milestone B acquisition phase. *Defense Acquisition Research Journal*, 20(3), 283–308.

Foreman, J. D. (2007). *Predicting the effect of longitudinal variables on cost and schedule performance* (Master's thesis). Defense Technical Information Center (DTIC) (ADA463520), Fort Belvoir, VA. http://www.dtic.mil/get-tr-doc/pdf?AD= ADA463520

Genest, D., & White, E. (2005). Predicting RDT&E Cost Growth. *The Journal of Cost Analysis & Management*, 7(1), 1–12.

Harmon, B. R. (2012, September). *The limits of competition in defense acquisition*. Paper presented at the Defense Acquisition University Research Symposium, Fort Belvoir, VA. http://www.dau.mil/Research/symposiumdocs/Harmon-Near substitutes paper.pdf

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70. https://www.ime.usp.br/~abe/lista/pdf4R8xPVzCnX.pdf

Jimenez, C. A., White, E. D., Brown, G. E., Ritschel, J. D., Lucas, B. M., & Seibel, M. J. (2016). Using Pre-Milestone B Data to Predict Schedule Duration for Defense Acquisition Programs. *Journal of Cost Analysis and Parametrics,* 9(2), 112–126. http://doi.org/10.1080/1941658X.2016.1201024

Lucas, B. M. (2004). *Creating Cost Growth Models for the Engineering and Manufacturing Development Phase of Acquisition Using Logistic and Multiple Regression*. (Master's thesis). Defense Technical Information Center (DTIC) (ADA422915), Fort Belvoir, VA. http://www.dtic.mil/get-tr-doc/pdf?AD= ADA422915

Neter, J., Kutner, M. H., Nachtseim, C. J., & Wasserman, W. (1996). *Applied linear statistical models* (4th ed.). Boston, MA: McGraw-Hill Companies.

Sipple, V., White, E., & Greiner, M. (2004). Using Logistic and Multiple Regression to Estimate Engineering Cost Risk. *The Journal of Cost Analysis & Management*, 6(1), 67–79.

Tisdel, J. E. (2006). *Small sample confidence intervals in log space back-transformed from normal space*. (Master's thesis). Defense Technical Information Center (DTIC) (ADA450276), Fort Belvoir, VA. Thesis. Air Force Institute of Technology, 2006. http://www.dtic.mil/get-tr-doc/pdf?AD= ADA450276

Appendix A: Implementation of Will-Cost and Should-Cost Management

DEPARTMENT OF THE AIR FORCE
WASHINGTON DC

JUN 15 2011

MEMORANDUM FOR SEE DISTRIBUTION

SUBJECT: Implementation of Will-Cost and Should-Cost Management

In order to gain greater efficiency and productivity in Defense spending, the Under Secretary of Defense for Acquisition, Technology & Logistics (USD(AT&L)) has directed the Military Departments and Directors of Defense Agencies to implement Will-Cost and Should-Cost management for all Acquisition Category (ACAT) I, II, and III programs. Dr. Carter, USD (AT&L), is challenging program managers to drive productivity improvements into their programs during contract negotiation and program execution by conducting Should-Cost analysis. This analysis goes beyond the Federal Acquisition Regulation/Defense Federal Acquisition Regulation Supplement (FAR/DFARS) Should-Cost reviews. FAR/DFARS Should-Cost reviews set realistic objectives for negotiating the immediate contract. The Should-Cost estimate as defined in this implementation memorandum is much broader in definition, covering all government and contract program costs throughout the entire life-cycle. SAF/AQ and SAF/FM fully support the implementation of Will-Cost and Should-Cost management and expect the Air Force acquisition community to embrace the concepts and adjust our management processes immediately.

The Department will continue to set program budget baselines using non-advocate Will-Cost estimates. Air Force guidance and instruction (e.g., AFPD 65-5 and AFI 65-508) describe specific requirements for non-advocate Will-Cost estimates or Service Cost Positions in support of ACAT I milestone decisions. However, the same level of rigor and attention is currently not required for ACAT II and III programs even though they account for about 48 percent of the Air Force acquisition budget. To ensure we exercise the same discipline for these programs that we do for our ACAT I programs, all ACAT II and III programs identified on the Acquisition Master List will present Will-Cost estimates at milestone decisions that have been approved by the appropriate product or logistics center financial management cost estimating organization (FMC). As with ACAT I programs, the non-advocate Will-Cost estimate will be used as the basis for all budgeting and programming decisions. All metrics and reporting external to the department will be based on the Will-Cost estimate.

Program managers must begin to drive leanness into their programs by establishing Should-Cost estimates at major milestone decisions. The Should-Cost estimate is an internal management tool for incentivizing performance to target, and is, therefore, not to be used for budgeting, programming, or reporting outside the department. Therefore, Should-Cost estimate documentation must be marked and treated as For Official Use Only. We recognize program managers have concerns about providing estimates that are lower than the budget, since DoD culture tends to use programming and budgeting to incentivize achievement. That is not the intent of this initiative. Will-Cost estimates are the official program position for budgeting, programming, and reporting.

Program managers are responsible for developing Should-Cost estimates. They should ensure cross-functional involvement in the development of the Should-Cost estimate and they can seek assistance from outside organizations (e.g., the Air Force Cost Analysis Agency or Defense Contract

Management Agency) throughout the development process. This effort does not necessarily require large teams to perform detailed bottoms-up assessments on every ACAT I, II, and III program. In some cases, this level of detailed analysis is extremely beneficial and desired, but we expect Program Executive Officers (PEOs), Designated Acquisition Officials (DAOs), and program managers to consider resources required versus potential benefits to determine the best approach. At a minimum, program managers are expected to identify specific discrete measurable items or initiatives that achieve savings against the Will-Cost estimate.

In accordance with USD AT&L direction, program managers for ACAT I, II and III programs identified on the Acquisition Master List will present Should-Cost estimates at their next major milestone. The Milestone Decision Authority (MDA) will approve all Should-Cost estimates and will expect program managers to manage, report, and track to these estimates. We will provide an annual report to OUSD (AT&L)/ARA on our progress. By 1 Jul 2011, PEOs/DAOs will submit a prioritized plan and timeline for completing Should-Cost estimates on all their ACAT I, II, and III programs not scheduled for a major milestone review in 2011. We recognize a waiver for some of these requirements may make sense. USD(AT&L) will consider and approve waivers for ACAT ID and IAM programs. SAF/AQ and SAF/FM will consider and approve waivers for all ACAT IC/IAC programs. The PEOs/DAOs and product/logistic center FM leads will approve waivers for ACAT II and III programs.

The following Air Force programs have been designated as pilots: JSF (F-35), Global Hawk Blocks 30 & 40, Evolved Expendable Launch Vehicle (EELV), Space Based Infrared System (SBIRS), and Advanced Extremely High Frequency (AEHF) Satellite System. These programs will be the first to actually have funds distributed based on Should-Cost execution baselines. The difference between the funds distributed and the program budget baseline will be held at the Service level. SAF/AQ and SAF/FM will jointly be the decision authority for release of these funds. We will need to capture lessons learned from each of these programs and share them with OSD and the other Services.

The attachment provides additional guidance and clarifies terms, procedures, and reporting requirements associated with this initiative. The guidance will be updated and codified in policy as USD(AT&L) and the Services/Components gain experience with Will-Cost and Should-Cost management. The POCs for this issue are Ms. Ranae Woods, AFCAA/TD, 703-604-0400, ranae.woods@us.af.mil and Mr. Bob Martin, SAF/AECO, 703-588-7177, robert.martin@pentagon.af.mil.

Jamie M. Morin
Assistant Secretary of the Air Force
(Financial Management and Comptroller)

David M. Van Buren
Air Force Service Acquisition Executive

Appendix B: Predictor Variables Investigated in this Paper

- *MS A to MS B Duration (Months) – Continuous Variable*
  - This variable indicates the total time it took in months for a program to complete MS A to MS B according to the last SAR date. In this variable we are only concerned with actual schedule duration data available to the cost estimator at the time of MS B/EMD start.

- *Quantity Expected at MS B – Continuous Variable*
  - This variable indicates the estimate of total quantity of weapons systems that were expected to be produced at MS B at the time of the last SAR date.

- *RDT&E Funding ($M) at MS B Start (BY17) – Continuous Variable*
  - This variable is based on raw total RDT&E dollars (in millions) that were allocated to the program prior to MS B. The dollars were all standardized into the base year when the research began (BY17).

- *(Projected) % of RDT&E Funding at MS B Start (BY17) – Continuous Variable*
  - This variable is based on the percent of available RDT&E dollars allocated to the program before, and up to the start of, MS B. While this variable is based on a percentage, the dollars that this percentage was derived from were all standardized into the base year when the research began (BY17).

- *(Projected) Total Program Acquisition Cost (BY17) – Continuous Variable*
  - This variable is the total projected acquisition costs, from MS B to IOC, estimated at MS B or the earliest available program SAR. It serves to identify how large a program is projected to be in terms of cost.

- *Modification – Binary Variable*
  - This variable is identifies programs whose existence serves as a modification to a pre-existing weapons system. If a weapons system is a modification, it does not necessarily mean it will not have pre-MS B data associated with it. Every program is different and, therefore, it cannot be assumed that a modification will automatically start at MS B.

- *Prototype – Binary Variable*
  - This variable identifies programs that create a prototype, or prototypes, of a weapons system before production of that weapons system begins. More than one type of prototype for a weapons system can be created in a given program.

- *Concurrency Planned – Binary Variable*
  - This variable addresses planned concurrency in a given program prior to MS B. Concurrency is the proportion of RDT&E dollars that are authorized

during the same years that Procurement appropriations are authorized. The planned level of concurrency forces managers to make decisions that can lead to [schedule] growth if either too much or too little concurrency is accepted for a given program (Birchler, Christle, & Groo, 2011, p. 246).

- *1985 or Later for MS B Start – Binary Variable*
  - This variable accounts for a time series trend of programs that started their MS B in 1985 or later. It is shown that programs which began development during 1985 or later (considered "contemporary") expend a greater percentage of obligations by their schedule midpoint than the earlier pre-1985 programs. We attribute this difference to the President's Blue Ribbon Commission on Defense (commonly called the Packard Commission) and the subsequent acquisition reforms.

- *Air Force – Binary Variable*
  - This variable identifies if the lead service on the program was the United States Air Force.

- *Navy – Binary Variable*
  - This variable identifies if the lead service on the program was the United States Navy.

- *Army – Binary Variable*
  - This variable identifies if the lead service on the program was the United States Army.

- *Marine Corps – Binary Variable*
  - This variable identifies if the lead service on the program was the United States Marine Corps.

- *Fixed Wing – Binary Variable*
  - This variable identifies if the weapons system program is a fixed wing aircraft program, regardless of service it is associated with. The criterion to qualify as a fixed wing aircraft is for that weapons system to maintain flight via fixed wings versus rotary wing flight.

- *Fighter Program – Binary Variable*
  - This variable identifies if the weapons system program is a fighter program, or close variation thereof, regardless of service it is associated with.

- *Bomber Program – Binary Variable*
  - This variable identifies if the weapons system program is a bomber program, or close variation thereof, regardless of service it is associated with.

- *Helo Program – Binary Variable*
    - This variable identifies if the weapons system program is a helicopter program, or close variation thereof, regardless of service it is associated with.

- *Cargo Plane Program – Binary Variable*
    - This variable identifies if the weapons system program is a cargo plane program, or close variation thereof, regardless of service it is associated with.

- *Tanker Program – Binary Variable*
    - This variable identifies if the weapons system program is a tanker plane program, or close variation thereof, regardless of service it is associated with.

- *Electronic Warfare Program – Binary Variable*
    - This variable identifies if the weapons system program is an electronic warfare program, or close variation thereof, regardless of service it is associated with. An electronic warfare program, as not to be confused with an electronic system program, differs greatly in its main function(s). A description from Lockheed Martin makes the distinction that it involves the ability to use the electromagnetic spectrum – signals such as radio, infrared or radar – to sense, protect, and communicate. At the same time, it can be used to deny adversaries the ability to either disrupt or use these signals (Electronic Warfare).

- *Trainer Plane Program – Binary Variable*
    - This variable identifies if the weapons system program is a trainer plane program, or close variation thereof, regardless of service it is associated with.

- *Missile Program – Binary Variable*
    - This variable identifies if the weapons system program is a missile program, or close variation thereof, regardless of service it is associated with.

- *Electronic System Program – Binary Variable*
    - This variable identifies if the weapons system program is an electronic system program, or close variation thereof, regardless of service it is associated with. This differs greatly from the previously described electronic warfare variable in that electronic systems programs are principally concerned with the electronic user interface of a system, avionics controls, or other similar applications that primarily support the electronic usability of a system, or system of systems.

- *Submarine Program – Binary Variable*
    - This variable identifies if the weapons system program is a submarine

program, or close variation thereof, regardless of service it is associated with.

- *Ship Program – Binary Variable*
  - This variable identifies if the weapons system program is a surface ship program, or close variation thereof, regardless of service it is associated with.

- *Satellite Program – Binary Variable*
  - This variable identifies if the weapons system program is a satellite program, or close variation thereof, regardless of service it is associated with.

- *ACAT I – Binary Variable*
  - This variable indicates if the program is an ACAT I program. This is significant in that ACAT I programs deal with a much larger dollar amount and thus are more susceptible to cost and schedule growth by way of their large-scale and complexity efforts.

- *(Projected) MS C to IOC Duration (Months) – Continuous Variable*
  - This variable indicates the total estimated time for a program to meet IOC from MS C according to the earliest available SAR estimate. This variable has been found to be predictive of cost growth in the programs studied by Foreman (2007). With this variable, we are concerned with giving the cost estimator the ability to enter in the projected duration, in months, of the gap between MS C and IOC to predict program cost.

- *(Projected) MS C Slip – Binary Variable*
  - This variable indicates whether the program projected date for meeting IOC extends past the initial estimate. Foreman (2007) has found that a slip in MS C is indicative of program cost growth in past research.

- *No MS A Date – Binary Variable*
  - This variable identifies whether a program did not contain a MS A date in the schedule portion of the SAR, but did include funding at least one year prior to MS B. This is used to identify these programs and test that they are not statistically different from the other programs and is not used in a predictive capacity.

- *Small Program – Binary Variable*
  - This variable identifies whether a program's projected total acquisition costs (RDT&E and Procurement) are below $3B. This value is determined from analyzing the histogram of the (projected) total program acquisition costs of the programs in our study and coincides closely with the 25% value.

- *Medium Program – Binary Variable*
    - This variable identifies whether a program's projected total acquisition costs (RDT&E and Procurement) are above $3B but below $7B. This value is determined from analyzing the histogram of the (projected) total program acquisition costs of the programs in our study and coincides closely with the 25% to 50% range.

- *Large Program – Binary Variable*
    - This variable identifies whether a program's projected total acquisition costs (RDT&E and Procurement) are above $7B but below $17.5B. This value is determined from analyzing the histogram of the (projected) total program acquisition costs of the programs in our study and coincides closely with the 50% to 75% range.

- *Extra Large Program – Binary Variable*
    - This variable identifies whether a program's projected total acquisition costs (RDT&E and Procurement) are above $17.5B. This value is determined from analyzing the histogram of the (projected) total program acquisition costs of the programs in our study and coincides with the 75% value.

- *(Projected) % Complete at MS B Start – Continuous Variable*
    - This variable is motivated by the % RDT&E variable and serves to project the percent that a program is complete, to IOC, when MS B occurs. It is calculated by dividing the projected duration from MS B to IOC by the sum of duration from MS A to IOC and projected duration from MS B to IOC. This serves to indicate where the program managers believe the program is in terms of schedule completeness. It could indicate program maturity level.

Appendix C: List of Programs Used in the Research Database

Number	Program	Number	Program
1	A-10	38	SSN 774 (Virginia Class Sub)
2	AWACS	39	T-45TS
3	C-17	40	UGM-109 Tomahawk
4	F-22	41	SSBN 726 SUB
5	AH-64	42	AGM-114A Hellfire Missile
6	B-1B Computer Upgrade	43	OH-58D Helicopter
7	C-5 RERP	44	AAWS-M Javelin
8	F-15	45	SSN 21 Sub
9	B-1B JDAM	46	AWACS Blk 40-50 Upgrade
10	KC-135R	47	B-2 EHF Inc 1
11	FA-18 A/B	48	C-5 AMP
12	AV-8B Harrier	49	MQ-9 Reaper
13	S-3A	50	AH-64E Remanufacture
14	P-8 Poseidon	51	ATACMS-APAM
15	V-22 Osprey	52	CH-47F
16	E-2C Hawkeye	53	CSSCS (ATCCS)
17	F-35 JSF	54	Longbow Apache (AH-64D)
18	CH-47D Chinook	55	UH-60M Blackhawk
19	E-8A JSTARS	56	AESA
20	AGM-65A Missile	57	AGM-88E AARGM
21	ALCM Missile	58	CEC
22	AMRAAM Missile	59	E-2D AHE
23	JASSM Missile	60	JSOW
24	JDAM	61	LCS
25	JPATS T-6A	62	LHD-1
26	OTH-B	63	MH-60R
27	LGM-118 Peacekeeper	64	MH-60S
28	GBU-39 SDB-I	65	Strategic Sealift
29	National Aerospace System	66	Trident II
30	AGM-88 HARM	67	EA-6B ICAP III
31	AIM-9X Block 1	68	JSIPS (CIGS)
32	AN/BSY-1	69	NAS
33	COBRA Judy Replacement	70	AFATDS (ATCCS)
34	Harpoon Missile	71	AEHF
35	NMT	72	EELV
36	SH-60B	73	WGS
37	UGM-96A Trident I Missile		

V. Estimating the Likelihood of a Defense Acquisition Program Staying within Cost and Schedule Bounds

Abstract

Program managers use prior experience to spot potential programmatic pitfalls and areas of concern. Augmenting this experience with an empirical procedure, we present a method to estimate the likelihood of a program exceeding two important schedule and cost thresholds: 1) under 15 percent over the initial cost estimate from Milestone (MS) B to Initial Operating Capability (IOC), and 2) 15 percent under the initial length (in months) between MS B and IOC; the second bound being under 25% respectively with respect to cost growth and schedule slippage. For our analysis, we use 49 Department of Defense programs. Using logistic regression and odds ratios, we generally find that electronic system programs, extremely large programs (exceeding \$17.5B in Base Year 2017 dollars), programs procuring smaller quantities of units, and programs with shorter schedules (less time from MS A to MS B and projected time from MS B to IOC) experience smaller percentages of cost growth and schedule slippage.

Estimating the Likelihood of a Defense Acquisition Program Staying within Cost and Schedule Bounds

Introduction

In today's fiscally constrained environment, we should use every tool at our disposal to contain cost growth and schedule slippage. As good stewards of the taxpayer, it is our duty to ensure that Department of Defense (DoD) programs are fielded on time and on budget. This also includes being aware of program characteristics that may lead to future cost growth and/or schedule slippage. To investigate this, we employ a statistical technique that is often adopted in the biostatistical community–logistic regression. Using this technique, we identify cost and schedule variables that may indicate a program will experience significant cost growth and/or schedule slippage. Specifically, we consider cost and schedule growth levels consistent with the Nunn-McCurdy significant and critical breach thresholds since these percentages have been identified by leadership as being growth above and beyond anything that can be considered acceptable.

With this in mind, we categorize defense acquisition programs based on their cost and schedule performance at the time they meet Initial Operating Capability (IOC) versus what they estimated at Milestone (MS) B. To this end, we consider a program to be "good" if they are within a specified percentage of their estimated cost and schedule and "bad" if they are not. The intent of our research is to ascertain what factors may be statistically significant in predicting the probability at MS B that a DoD acquisition program will fall into either category. We use the Nunn-McCurdy threshold of 15% over the current baseline as our cutoff point for "significant" growth (Schwartz & O'Connor, 2016). We replicate this process for increases of 25% and consider those a "critical' overrun.

Background and Database

To the best of our knowledge, the literature appears to be scant with DoD studies which simultaneously analyze program cost and schedule performance. One exception is a National Aeronautics and Space Administration (NASA) study (Burgess & Krause, 2014) that looked at the interaction between the phasing estimating relationship (PER), the cost estimating relationship (CER), and the schedule estimating relationship (SER). The CER is the total program cost from the System Requirements Review (SRR), a date that occurs prior to MS B, through launch; while the SER is the time, in months, from SRR to launch. Given these cost and schedule estimates, the PER relays the annual funding profile for the program. They used historical data from 37 NASA programs for their study and developed multiple regression models to analyze these relationships.

From their analysis, they developed a set of tools to give decision makers the ability to quantify trade-offs between cost, schedule, and phasing in their program. Additionally, Burgess and Krause's (2014) analysis allowed the program manager to conduct a programmatic "health assessment" in which the estimating relationships are analyzed to determine if they fall within a standard deviation of the mean observed historical value. Our research deviates from theirs in that we analyze categorical indicators for programs that fall under or over the 15% and 25% current and original baseline thresholds. We aim to describe what indicators may correlate to a program being deemed "good" or "bad" in the future based upon characteristics at MS B.

Before considering candidate variables that may be indicative of a DoD acquisition program being “good” or “bad”, we assume MS I, II, and III to be equivalent to MS A, B, and C, respectively. This is based on their respective definitions and that the naming convention has simply changed over time without a tangible change in definition or substance as noted by Harmon (2012) and Jimenez, White, Brown, Ritschel, Lucas, and Seibel (2016).

From Burgess and Krause (2014), we find two predictive results: 1) longer duration from SRR to Preliminary Design Review (PDR) suggests increased likelihood of program schedule lengthening, and 2) higher percent of new designs appear to increase likelihood of increased cost in acquisition programs. Jimenez et al. (2016) determined the following variables are statistically significant for predicting increased schedule duration: whether a program is a new effort or modification to an existing program, the amount of raw funding (adjusted for inflation) prior to MS B for a program, and the percentage of total RDT&E (Research Development Test and Evaluation) funding profile allocated at MS B. They also suggest that information obtained prior to MS B data, such as length between MS A and MS B, may also prove useful in predicting a program’s length.

A study conducted by Deitz, Eveleigh, Holzer, and Sarkani (2013) examined the importance of developing a robust Analysis of Alternatives (AOA) prior to MS B and the effects it may have on program success. The most important finding of their research is while only 10% of a program’s lifecycle cost are invested prior to MS B, 70% of a program’s lifecycle costs are committed to by MS B (Deitz et al., 2013). Similar to Jimenez et al. (2016), this suggests pre-MS B data is very important to predicting program outcomes. Unfortunately not many programs have pre-MS B data since a DoD acquisition program does not officially begin until MS B. This data limitation will be discussed later.

Based on our literature review, we gain insight into what variables may prove useful to predict a program's cost and schedule as well as the likelihood cost growth and/or schedule slippage. Additionally, we use the research to develop inclusion and exclusion criteria for our database. Programs in our study must meet the following three criteria: 1) be unclassified, 2) have Selected Acquisition Report (SAR) data in the Defense Acquisition Management Information Retrieval (DAMIR) system, and 3) are designated as a Major Defense Acquisition Program (MDAP) or pre-MDAP within DAMIR.

Conversely, there are four exclusion criteria for this research. The first requirement is the program SAR must contain a MS A date or funding in the funding profile at least one year prior to MS B (indicates the year in which MS A may have occurred). This requirement is due to the pre-MS B data being found predictive in the past research we studied. Unfortunately, this requirement also results in a great deal of programs being ineligible for inclusion because a lack of reporting requirements prior to MS B since the program is not official until meeting MS B.

The second exclusion criteria is that the program SAR must contain a MS B date and corresponding funding information. This again pertains to the necessity of containing pre-MS B data as a means to build a predictive model. Without the MS B date and funding information, we are unable to ascertain the duration of MS A or the funding spent up to MS B. Additionally, we are unable to calculate the projected funding needed to reach IOC or the projected duration of MS B to IOC.

The third exclusion criteria is the program SAR must contain an IOC date that occurred prior to the last reported SAR; this indicates that the program is complete up to IOC. This is important to our research as it gives us a termination point to estimate and ensures we are not using projected values as actuals in our model. IOC is also important as it signifies the point in time when the user community initially benefits from the investment in the program.

The fourth exclusion requirement is that the program must contain a SAR within one year of reaching MS B. This requirement allows us to ascertain what the program's cost and schedule estimate was at MS B and if the actual cost and schedule from MS B to IOC is within 15% or 25% of this estimate. Note, we allow one year from the time MS B occurs because the program may not have been required to report a SAR at the time MS B occurred. Table 1 summarizes the inclusion and exclusion criteria. Based on these criteria, we use 49 programs in our analysis—the specific programs are listed in the Appendix.

Table 1: Program Inclusion Table

Inclusion/Exclusion Criteria	Program Included	Programs Removed	Program Count
Jimenez et al. (2016) Database	56		56
DAMIR Query (MDAP/Pre-MDAP)	187		243
Doubled counted from Jimenez et al. (2016) Database		29	214
IOC Occurs after Last SAR		61	153
Missing Milestones A or B		74	79
No SAR within 1 year of MS B		24	55
Missing IOC		4	51
Classified		2	49
Final Number			49

For each of the 49 DoD acquisition programs in our database, we use two SARs. For the response variables, we use the last reported SAR for each program to gather the actual cost and schedule duration for each program from MS B to IOC. For the candidate explanatory variables, we use the SAR from the year in which MS B occurred or, if this is unavailable, the SAR within one year of MS B. The cost and schedule estimate from MS B to IOC in this SAR becomes the current estimate with respect to measuring cost growth and schedule slippage. The cost growth percentage is calculated as *Current Cost Estimate at MS B – True Cost from MS B to IOC / Current Cost Estimate at MS B.* A similar calculation is computed for schedule.

Methodology

As noted earlier, the purpose of this article is to identify predictor variables that may determine the likelihood that a DoD acquisition program will experience cost growth and/or schedule slippage above certain thresholds. To account for the fact that programs change, possibly due to forces outside of the program manager's control, we employ two separate threshold values. The first is the 15% threshold above the current estimate (both cost and schedule) from MS B to IOC established at MS B. The second threshold is set at 25%. We chose these overrun thresholds based upon the significant (15%) and critical (25%) definition of Nunn-McCurdy breaches. For 30 years, the Nunn-McCurdy Act (10 U.S.C. §2433) has served as one of the principal mechanisms for notifying Congress of cost overruns in MDAPs—a MDAP is defined as a program estimated to have research and development costs greater than $480M or procurement costs greater than $2.79B (in FY2014 constant dollars) (Schwartz & O'Connor, 2016).

As previously mentioned, we investigate both the total acquisition cost and duration from MS B to IOC. All costs in our models are in Base Year 2017 (BY17) dollars using the 2016 OSD inflation indices, which prevents inflation from influencing our model. For the 15% and 25% response categories, we assign each of our 49 programs in the database to one of four mutually exclusive categories: "Good/Good", "Good/Bad", "Bad/Good", and "Bad/Bad". A program is considered “good” if the final cost growth (or schedule slippage) from MS B to IOC is less than the chosen overrun threshold; a program is considered “bad” if it equals or exceeds the overrun threshold.

Initially, we aimed to identify variables that may predict which of the four categories a DoD program might fall into at MS B; however, the limited sample sizes for “Good/Bad” and “Bad/Good” prevented this. Combining these groups only resulted in nine programs with a significant overrun and six programs with a critical overrun. Since this combined category lacked the requisite statistical power to conduct a logistic regression analysis, we only focus on the “Good/Good” and “Bad/Bad” categories for both the 15% and 25% thresholds. These designations are listed in the Appendix for each program in our database.

To build our initial logistic regression model, we use a mixed stepwise approach to identify the most predictive variables; a 0.1 level of significance was selected for the entry and exit criteria due to the exploratory nature of our work. For the finalized model, the resultant predictor variables from the stepwise procedure must meet the overall model Type I error of 0.1 and require each variable to be significant according to the Holm-Bonferroni criteria (Holm, 1979). We use JMP® Pro 12 for all statistical analysis performed in this article.

A logistic regression model predicts the probability of a program identifying with a particular group by way of the following equation:

$$y = \frac{e^{f(x)}}{e^{f(x)} + 1} \tag{1}$$

where y is a binary variable indicating a program’s group, e is the natural exponent function, and $f(x)$ is considered the logit or log-odds function (Gaudard, Ramsey, & Stephens, 2006) and can be written in the form:

$$f(x) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \qquad (2)$$

Equation (1) represents an *s*-shaped curve (White, Sipple, & Greiner, 2004) whose values range from 0 to 1 (probability).

The *X* variables in (2) typify the standard explanatory variables used in linear regression, however, the β coefficients do not represent the mean change in the response. Instead, e^{β_i} represents the odds ratio (OR) of a particular program in our database belonging to either "Good/Good" or "Bad/Bad" when the *X* variables are dichotomous (i.e., $X_i = 1$ when a characteristic is present or $X_i = 0$ when a characteristic is not present). Continuous explanatory variables do not possess this easy interpretation of ORs because there is no natural baseline group to compare. Therefore, all explanatory variables have been converted to this dichotomous setting.

For categorical variables, this transformation is straight-forward. For example, a dummy variable might be coded a '1' if the program is an Air Force acquisition program, '0' otherwise (i.e., is an Army, Marine, or Navy program). For continuous data, we discretize (i.e., create categorical groupings) by utilizing histograms to determine potential break points in the data. These break points often coincide with the quartiles (25th percentage, 50th percentage, or 75th percentage) of the histograms.

We use two metrics to quantify the predictive capability of our logistic regression models. The first metric is the Area Under the Receiver Operating Characteristic Curve (AUC). The AUC indicates the sorting efficiency of a model with a value of 0.5 indicating merely random chance and a value of 1.0 indicating perfect prediction capabilities (Gaudard, Ramsey, & Stephens, 2006). The AUC is a single measure of the overall discrimination ability of a test. In general, "an AUC that is greater than 0.8 suggests that the diagnostics test has good discriminatory power" (McPherson & Pincus, 2016: 80). Since we have such a small subset of data for each group, it is infeasible to set aside a 20% validation pool. Given this limitation, we use a technique called bootstrapping (Efron & Tibshirani, 1994) to present a 90% confidence interval for the AUC value for each logistic regression model; these intervals provide the user predictive limitations of the model.

The second metric to demonstrate the utility of our logistic regression models is the OR for each explanatory variable and its corresponding confidence bound (either the lower or upper value in the confidence interval that is closet to the value of 1). An OR equal to 1 indicates the explanatory variable does not effect the odds of a program belonging to either the "Good/Good" or "Bad/Bad" category. An OR > 1 implies a higher odds of a program belonging to the "Good/Good" category, while an OR < 1 suggests a lower odds of belonging to the "Good/Good" category (Szumilas, 2010). With respect to the confidence interval of an odds ratio, either the lower or upper confidence bound is used to estimate the precision of the OR. In practice, this bound is often used as a proxy for the presence of statistical significance if it does not overlap the null value (e.g., OR = 1) (Szumilas, 2010).

Lastly, to prevent model extrapolation, the ranges of the continuous independent variables over which the models are useful must be consistent with the bounds of the programs used in our analysis. Using the models outside these ranges may invalidate the results. Only three continuous explanatory variables proved statistically significant in our models. For projected duration from MS B to IOC the range is 30 to 109 months. For projected percent complete at MS B, the range is 15% to 70%. For the duration from MS A to MS B, the range is 13 to 125 months.

Results

The following subsections illustrate the logistical models derived from the stepwise procedure along with an explanation of each significant explanatory variable. The first subsection highlights the results regarding the "Good/Good" and "Bad/Bad" groups for the 15% overrun threshold (*Significant*), while the second subsection highlights the results for the "Good/Good" and "Bad/Bad" groupings for the 25% overrun threshold (*Critical*).

Significant Overrun

For this analysis, 15 programs (approximately 31% of our database) fall in the "Good/Good" group and 25 (approximately 51% of our database) programs are in the "Bad/Bad" group. Table 2 summarizes the logistic model and associated predictive explanatory variables for determining the likelihood of a DoD acquisition program experiencing less than 15% cost and schedule growth from MS B to IOC. The model has an AUC of 0.88 suggesting good model discrimination. All of the estimated ORs and their associated confidence bounds are well above or below 1.

Table 2: Significant predictor variables for determining the likelihood a program will experience cost growth and schedule slippage less than 15 percent. Numbers rounded to two significant digits. AUC = 0.88 with a 90% bootstrapped confidence interval (1000 samples) of (0.84, 0.98). The family-wise error rate for the independent variables is 0.10.

Variable	Estimate	Odds Ratio	Odds Ratio Bound	Chi-Square	P-Value
Intercept	–3.32	N/A	N/A	3.20	0.0735
Projected MS B to IOC <= 58 months	3.63	37.83	5.55	7.37	0.0066
Extra Large Program	3.37	29.13	4.58	7.01	0.0081
Electronic System Program	3.27	26.37	3.64	6.09	0.0136
Projected % Complete at MS B <= 35%	3.32	27.76	3.98	5.99	0.0144
MDAP	–3.34	0.036	0.49	3.75	0.0529

The Electronic System Program variable indicates if the DoD acquisition program is an electronic user interface system, avionics control system, radio network system, or similar electronic system. The OR suggests that such systems typically display cost growth and schedule slippage less than 15%. This appears to be in keeping with Bolten, Leonard, Arena, Younossi, and Sollinger (2008) who also concluded that electronic systems are historically cheaper.

For the Projected MS B to IOC Duration <= 58 Months explanatory variable, this finding suggests that acquisition programs whose projected MS B to IOC duration is equal to or less than 58 months (or approximately 5 years) typically display cost growth and schedule slippage less than 15%. We theorize this may be indicative of relatively shorter-scoped programs whose technology may be relatively more mature.

The Extra Large Program explanatory variable suggests that acquisition programs with a high cost (greater than \$17.5B BY17 dollars in total project acquisition cost) typically experience cost growth and schedule slippage less than 15%. This is logically expected given larger programs do not have the flexibility of having sizeable overruns given the sheer amount of dollars involved before DoD oversight and/or Congressional reviews intervene and possibly cancel the program. Thus, we treat Extra Large Program as more of a covariate than a traditional explanatory variable.

The programs identified as MDAP in our database tend to suggest that this explanatory variable will lead to cost growth and schedule slippage greater than or equal to 15%. We believe this to be an artifact of our database due to the large number of programs that identify as MDAP (45 of 49, or 92%) and the fact that all the programs in the "Bad/Bad" group are identified by this variable. It is also noteworthy that three of the four programs not identified as MDAP are in the "Good/Good" group.

Finally, the Projected % Complete at MS B <= 35% variable (calculated as the actual time from MS A to MS B divided by the sum of the actual time from MS A to MS B and projected time from MS B to IOC) is statistically significant; this result suggests that programs that spend less time in the MS A-to-MS B phase relative to in comparison from MS A-to-IOC phase experience less cost growth and schedule slippage. This may be due to a high technology readiness level (TRL) early in the program's life or a lesser extent of new technology involved in the program. Such a conclusion is consistent with Dietz et al. (2013) who studied the pre-MS B process to identify cost estimating relationships associated with identified TRLs. Their findings indicate that programs with a higher TRL entering MS B experience smaller levels of cost growth.

Regarding the "Bad/Bad" group, Table 3 displays the logistic model and associated predictor variables for determining the likelihood of a DoD acquisition program's actual MS B-to-IOC cost and schedule exceeding its MS B estimate by 15% or more. The model has an AUC of 0.85 suggesting good model discrimination. All of the estimated ORs and their confidence bounds are well above or below 1.

Table 3: Significant predictor variables for determining the likelihood a program will experience cost growth and schedule slippage equal to or exceeding 15 percent. Numbers rounded to two significant digits. AUC = 0.85 with a 90% bootstrapped confidence interval (1000 samples) of (0.79, 0.95). The family-wise error rate for the independent variables is 0.10.

Variable	Estimate	Odds Ratio	Odds Ratio Bound	Chi-Square	P-Value
Intercept	1.41	N/A	N/A	2.65	0.1038
Extra Large Program	–4.60	0.01	0.09	8.89	0.0029
Electronic System Program	–3.15	0.04	0.26	6.74	0.0094
Aircraft	3.29	26.86	3.19	5.00	0.0254
RDT&E at MS B Start >= $272M	1.87	6.48	1.64	4.47	0.0346
Qty Expected at MS B <= 305	–1.95	0.14	0.62	3.98	0.0461

Similar to the "Good/Good" model, both the explanatory variables of Extra Large Program and Electronic System Program are statistically significant. However, both variables have negative parameter estimates (and thus ORs much smaller than 1), which indicates programs displaying these characteristics are much less likely to experience cost growth and schedule slippage equaling or exceeding 15%. This is consistent with our findings from the "Good/Good" group in Table 2.

The explanatory variable identifying a program as a fixed wing aircraft is statistically significant in predicting whether a program is more likely to experience cost growth and schedule slippage equaling or exceeding 15%. We believe this is due to the large and complex nature of these programs, especially given the modern aircraft programs, such as the F-22 and F-35, in our study.

Programs in our database that expect to procure less than or equal to 305 units at MS B tend to indicate that they are less likely to experience cost and schedule growth equaling or exceeding 15%. We believe that this may be due to uncertainty in estimating the variable costs for a program and the assembly line schedule. For example: if a program discovers that a variable cost is higher than estimated, the cost is multiplied by the number of units and the more units to be built, the higher the cost growth. The same can be said for underestimating the time a unit will take to assemble.

The last predictor variable associated with programs experiencing cost growth and schedule slippage equaling or exceeding 15% is for programs that spend greater than $272M in RDT&E by the start of MS B. We believe this may be indicative of programs with a low level of technological maturity thus requiring larger and more complex development prior to MS B. It could also indicate that a program is integrating many highly sophisticated components and the final design is complex in nature. As mentioned for the "Good/"Good" model, this is consistent with Dietz et al. (2013), who researched the pre-MS B process and found that a lack of maturity at MS B correlates with higher costs.

Critical Overrun

For this analysis, 20 programs (approximately 41% of our database) fall in the "Good/Good" group and 23 (approximately 47% of our database) programs are in the "Bad/Bad" group. Table 4 shows the logistic model and associated predictor variables for determining the likelihood that a DoD acquisition program's true MS B-to-IOC cost and schedule will be less than 25% larger than its MS B estimate. The model has an AUC of 0.84 suggesting good model discrimination. All of the estimated ORs and associated 90% confidence bounds are well above or below 1.

Table 4: Significant predictor variables for determining the likelihood a program will experience cost growth and schedule slippage less than 25 percent. Numbers rounded to two significant digits. AUC = 0.84 with a 90% bootstrapped confidence interval (1000 samples) of (0.78, 0.93). The family-wise error rate for the independent variables is 0.10.

Variable	Estimate	Odds Ratio	Odds Ratio Bound	Chi-Square	P-Value
Intercept	5.41	N/A	N/A	7.19	0.0073
Extra Large Program	3.34	28.25	5.01	7.58	0.0059
MDAP	–4.54	0.011	0.13	7.19	0.0073
MS A to MS B >= 28 Months	–2.99	0.05	0.27	6.33	0.0119
1985 or Later for MS B Start	–1.69	0.19	0.69	4.08	0.0434

With respect to previous results regarding MDAP and Extra Large Programs, we see similar results in this section. Extra Large Programs appear more likely to have cost growth and schedule slippage less than 25%, while MDAPs are less likely to have cost growth and schedule slippage under 25%.

For the MS A to MS B greater than or equal to 28 months explanatory variable, these programs appear less likely to experience cost growth and schedule slippage less than 25%. A possible explanation is that programs with relatively longer duration from MS A to MS B may indicate a program is relying upon complex technology that must be matured, which we believe is consistent with prior research conducted by Dietz et al. (2013).

The variable 1985 or Later for MS B Start indicates if a program is considered to be a part of the "modern" era of defense acquisition. These programs indicate that modern programs appear less likely to experience cost growth and schedule slippage less than 25%. This could be due to the increasing complexity of modern programs, which include the JSF and other highly complex systems, and that increased complexity drives cost and schedule. This is consistent with the work conducted by Jimenez et al. (2016), who found that these modern programs tend to have a longer schedule.

Regarding the "Bad/Bad" group, Table 5 displays the logistic model and associated explanatory variables for determining the likelihood of a DoD acquisition program having its true cost and schedule from MS B to IOC exceeding its MS B estimate by 25% or more. The model has an AUC of 0.79 suggesting fair to good model discrimination. All of the estimated ORs and associated confidence bounds are well above or below 1.

Table 5: Significant predictor variables for determining the likelihood a program will experience cost growth and schedule slippage equal to or exceeding 25 percent. Numbers rounded to two significant digits. AUC = 0.79 with a 90% bootstrapped confidence interval (1000 samples) of (0.70, 0.89). The family-wise error rate for the independent variables is 0.10.

Variable	Estimate	Odds Ratio	Odds Ratio Bound	Chi-Square	P-Value
Intercept	0.54	N/A	N/A	1.47	0.2253
Electronic System Program	–2.74	0.06	0.33	5.72	0.0168
Extra Large Program	–2.51	0.08	0.40	4.98	0.0257
Aircraft	2.10	8.19	1.52	3.20	0.0737

With respect to the explanatory variables of Extra Large Programs, Fixed Wing Aircraft, and Electronic System Programs, we see the same trends as we did in Tables 2–4; Extra Large Programs and Electronic System Programs are less likely to experience cost and schedule growth greater than 25%, while Fixed Wing Aircraft are more likely to experience cost growth and schedule slippage greater than or equal 25%. There are no additional significant variables for this model.

Discussion and Conclusion

In this article, we first investigate possible variables that statistically predict the likelihood of a DoD acquisition program experiencing cost growth and schedule slippage less than 15%. We also model the likelihood that a program would experience cost and schedule growth in excess of (or equal to) 15%. These percentage increases are measured with respect to the MS B-to-IOC estimates at MS B and the actual cost and schedule realized for MS B-to-IOC. We then replicate this process to determine which variables may be predictive if the threshold percentage increased from 15% to 25%. We chose these percentages based upon the significant and critical thresholds as defined by the Nunn-McCurdy act.

Overall, we determined the following five variables appear to be predictive factors for determining if a DoD acquisition program will experience less cost and schedule growth: Electronic System Programs, programs having a Projected MS B to IOC Duration less than (or equal to) 58 Months, Extra Large Programs, programs that expect to procure fewer than 305 units at the time of MS B, and programs with a Projected % Complete at MS B less than (or equal to) 35%. In contrast, MDAPs, Fixed Wing Aircraft, programs where the duration between MS A to MS B greater than (or equal to) 28 months, programs whose Projected % complete at MS B is greater than 38% appear, modern programs that enter MS B in 1985 or later, and programs that spend greater than, or equal to, $272M (FY17) of RDT&E funding by the beginning of MS B to be predictive that programs are likely to experience more cost growth and schedule slippage.

Our findings with respect to variables that incorporate the time between MS A and MS B are consistent with those of Dietz et al. (2013). These results suggest that programs with more technology uncertainty or immaturity at MS B have an increased likelihood of incurring higher cost growth and schedule slippage compared to more technologically mature programs. Additionally our findings with respect to Electronic Systems programs are supported by Bolten et al. (2008), though we do acknowledge that most of those programs in our database were both small in nature (under $3B BY17) and consisted primarily of modifications.

As with any statistical model, there are limitations to our logistic regression models. First, the database was created from SARs which may contain incomplete information. The models built are only as good as the data used to create them. There were multiple constraints on the data collecting process that hampered the ability to create a more robust database; foremost, the lack of availability of pre-MS B data limited the programs that could be included. Additionally, the search parameters in DAMIR may have unintentionally excluded programs which could have influenced the outcome of our analysis.

In order to gain insight on a program's potential for cost and schedule growth at such an early stage as MS B, we attempt to leverage the knowledge of the past to see where others have been. Our models may give program managers a glean at where they may be heading and highlight potentials pitfalls. This set of logistic regression models are designed to provide a tool for the DoD acquisition community to make strategic program health assessments. Practically, these models offer the potential to help portfolio managers decide where to allocate risk dollars.

Our research differs from prior research in that our database is expanded beyond the NASA only programs that were researched by Burgess and Krause (2014). Additionally, we utilize program characteristics across a large range of programs in order to develop logistic models that predict the probability of overrunning thresholds identified as being above acceptable levels of cost and schedule growth. No other research, to our knowledge, relates cost or schedule growth probability to overruning such important thresholds. The models may provide managers the ability to predict the possibility and severity of an overrun.

References

Bolten, J.G., Leonard, R.S., Arena, M.V., Younossi, O. & Sollinger, J.M. (2008). *Sources of weapon system cost growth: Analysis of 35 major defense acquisition programs* (MG 670). Santa Monica, CA: RAND Corporation.

Burgess, E., & Krause, C. (2014, August). *Integrated estimating relationships*. Paper presented at the NASA 2014 Cost Symposium, Hampton, VA. Retrieved from https://www.nasa.gov/sites/default/files/files/09_Burgess_NASA_Cost_Symposium_2014_IERs_Final_Tagged.pdf

Deitz, D., Eveleigh, T. J., Holzer, T. H., & Sarkani, S. (2013). Improving program success through systems engineering tools in the pre-milestone B acquisition phase. *Defense Acquisition Research Journal*, 20(3), 283–308.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Boca Raton, FL: Chapman and Hall/CRC Press.

Gaudard, M., Ramsey, P., & Stephens, M. (2006). *Interactive data mining and design of experiments: The JMP® partition and custom design platforms*. Cary, North Carolina: North Haven Group, LLC. http://islab.soe.uoguelph.ca/sareibi/PROJECTS_dr/GRAD_FUTURE_dr/docs/Interactive_DataMining.pdf

Harmon, B. R. (2012, September). *The limits of competition in defense acquisition*. Paper presented at the Defense Acquisition University Research Symposium, Fort Belvoir, VA. http://www.dau.mil/Research/symposiumdocs/Harmon-Near substitutes paper.pdf

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70. https://www.ime.usp.br/~abe/lista/pdf4R8xPVzCnX.pdf

Jimenez, C. A., White, E. D., Brown, G. E., Ritschel, J. D., Lucas, B. M., & Seibel, M. J. (2016). Using pre-milestone B data to predict schedule duration for defense acquisition programs. *Journal of Cost Analysis and Parametrics*, 9(2), 112–126. http://doi.org/10.1080/1941658X.2016.1201024

McPherson, R. A., & Pincus, M. R. (2016). *Henry's clinical diagnosis and management by laboratory methods* (23rd ed.). Amsterdam, Netherlands: Elsevier.

Schwartz, M., & O'Connor, C. V. (2016). *The Nunn-McCurdy act: Background, analysis, and issues for Congress* (Report No. R41293). Washington, DC.: Congressional Research Service.
Retrieved from https://fas.org/sgp/crs/natsec/R41293.pdf

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19(3), 227–229.

White, E. D., Sipple, V. P., & Greiner, M. A. (2004). Using logistic and multiple regression to estimate engineering cost risk, *Journal of Cost Analysis and Management*, 6(1), 67–79. http://www.tandfonline.com/doi/abs/10.1080/15411656.2004.10462248

Appendix: List of Programs and Their Respective Designations

Program	15% Group	25% Group	Program	15% Group	25% Group
A-10	3	3	COBRA Judy Replacement	3	2
C-17	3	3	Harpoon Missile	3	3
F-22	3	3	NMT	2	2
AH-64	3	3	SH-60B	3	3
B-1B Computer Upgrade	1	1	UGM-96A Trident I Missile	2	1
C-5 RERP	3	3	SSN 774 (Virginia Class Sub)	2	1
F-15	1	1	UGM-109 Tomahawk	1	1
B-1B JDAM	1	1	SSBN 726 SUB	3	3
FA-18 A/B	2	1	AGM-114A Hellfire Missile	3	3
AV-8B Harrier	1	1	OH-58D Helicopter	1	1
P-8 Poseidon	1	1	AAWS-M Javelin	3	2
V-22 Osprey	3	3	B-2 EHF Inc 1	1	1
F-35 JSF	3	3	AH-64E Remanufacture	2	2
CH-47D Chinook	2	1	CH-47F	3	3
E-8A JSTARS	3	3	UH-60M Blackhawk	3	3
ALCM Missile	3	3	AESA	1	1
AMRAAM Missile	1	1	AGM-88E AARGM	3	3
JASSM Missile	3	3	CEC	3	3
JDAM	1	1	E-2D AHE	3	3
JPATS T-6A	3	3	LCS	1	1
GBU-39 SDB-I	1	1	MH-60S	3	3
National Aerospace System	2	2	AEHF	3	3
AGM-88 HARM	2	2	EELV	1	1
AIM-9X Block 1	2	1	WGS	3	3
AN/BSY-1	1	1			

Code 1 implies cost growth and schedule slippage less than 15% (or 25%)
Code 2 implies either cost growth or schedule slippage less than 15% (or 25%) but not both
Code 3 implies cost growth and schedule slippage equal to or greater than 15% (25%)

VI. Conclusions and Recommendations

Chapter Overview

This chapter summarizes the findings in our research, from the cost regression model to the four logistic regression models. We begin by summarizing these findings before we discuss how our research questions have been answered. We then discuss the limitations of our research and conclude by discussing further research associated with ours.

Findings

Given that our research is broken into two parts, we discuss each of them in turn. We begin with the cost model analysis. This portion of our research focuses on predicting the median cost of a DoD program, similar to the 73 analyzed in our database, to give the program manager a means to impartially estimate the will-cost of a program, giving them a starting point to cross check the program estimate and a building block to perform the should-cost estimate. There are several findings that we consider important from this research.

The first finding we want to highlight is that MS C information, either MS C Slipping or projected duration from MS C to IOC, was not significant for predicting cost. This is contrary to the findings from the findings by Foreman (2007). This finding allowed us to include three satellite programs to our database, which would have been ineligible for inclusion otherwise due to their lack of MS C data.

Adding to this finding, the U.S. Government Accountability Office conducted a study on space programs and presented their findings before the U.S Senate on 11 May 2011. In their findings, they identify that space acquisition programs have historically experienced disproportionally large cost and schedule growth (GAO, 2011). This initially lead us to believe that these programs are not statistically similar to other DoD programs as the research conducted by Jimenez et al (2016). By including three space programs in our analysis and not finding that they were statistically different, we can call that conclusion into question. It must be noted, however, that we included only three space programs in a total database of 73 programs. Therefore, we may not have had the statistical power to detect any difference.

Of the five variables found significant in the cost model, one of the highest contributors to our model is the amount of RDT&E funding spent at the start of MS B. Interestingly, this is consistent with the findings by Jimenez et al. (2016) who found this variable to be predictive of a longer schedule. This leads us to believe that front loading a program most likely indicates a very technologically complex program which will logically have higher costs and schedule.

Shifting focus to the four logistic regression models, we are forced to remove additional programs from our study due to data restrictions driving the program count down to 49. This, foreseeably, lead to restrictions in terms of validating our analysis and giving us the statistical power to detect other potential predictor variables. Unfortunately, this means that we could not validate our models as is customary. This is unavoidable, and can be addressed by future research by expanding the database to include more programs.

Next, we highlight the internal consistency that these models demonstrate. Electronic System Programs, which tend to be Small Programs as well, consistently indicate that the program will not experience an overrun. Additionally, the Extra Large Programs indicate in every model that these programs are less likely to experience an overrun. These highlight certain program characteristics that suggest for further research to determine an underlying cause for why they indicate the potential for an overrun or not, as well as highlights the need for adjusting for program size in this research. We present and discuss conjecture on why these variables may indicate the potential for an overrun in Chapter V, the next step for future research is to determine why they indicate this.

Across the two portions of this research, we note many internal consistencies. The first consistency is that electronic system programs are relevant when predicting cost and predicting the probability of an overrun. As mentioned previously, these programs tend to be cheaper and are less likely to experience an overrun. These may be programs that inexperienced personnel should gain experience on without fear of large negative consequences. Another consistency we note is that pre-MS B data is important for programs to monitor and archive. This research and past research, as already mentioned, show that the inception of a program can have a drastic effect on how it performs. As such, more effort should be put towards monitoring and controlling this process.

Research Questions

The first question we set out to answer was: How can we use and build upon a previously created database to develop a mathematical model to predict the median cost of a program? In our research we leveraged the 56 program database built by Jimenez et al. (2016) and added 17 more programs and additional analytical variables to build a highly predictive model for the median cost from MS B to IOC in a DoD program. By finding these additional programs and predictive variables we are able to move the process of cost estimating forward into a data driven process.

Answering the second question: How can we identify program characteristics for significantly or critically overrunning either cost, schedule, or both given the current APB, at MS B through IOC, and predict the probability that a program will experience such overruns? In conducting this analysis, we are forced to include only 49 programs due to the requirement of having a program estimate within a year of MS B, which left us with a lack of power to detect the probability of only overrunning either cost or schedule, but not the other. We are able to give logistic regression models to program managers that identify several program characteristics for either significantly, or critically, overrunning both cost and schedule, or neither.

Limitations

There are several limitations to our findings that effect the applicability of this research to the field that we are able to identify. Foremost, the lack of available program data prior to MS B limited the inclusion of many programs. The ability of pre-MS B data to predict program cost dictates the use of this exclusion criteria, as this data has been found to be highly predictive, so the limitation cannot be avoided. By including 73 programs in the cost regression model, however, we have a good-sized data pool to draw significant inference.

Shifting to the logistic regression analysis of this thesis, the requirement to remove 24 programs due to the lack of an available program estimate within a year of MS B lead to further issues with our ability to identify predictive measures for cost and schedule overruns as outlined previously in this thesis. This analysis is affected to such an extent that we are unable to predict the programs that lie in the middle of our analysis and overrun either cost or schedule, but not both. By adding more programs to this analysis there may be enough data to identify delineating factors that bind the programs in this category. Additionally, this leaves us with too small of a data pool to validate any of our logistic models, leaving the models un-tested when presented with new data.

An additional limitation to our research is the fact that the analysis is only as good as the data that goes into it. The data is gathered through the program SARs, which may contain faulty or incomplete data. This is an accepted limitation due to the availability of program SARs and the lack of additional means to acquire the same program data. The missing data from the SARs lead to the inclusion of only 73 programs in the cost model, and further reduced the programs to 49 in the logistic regression models, thus limiting the ability for us to make inferences in general.

Lastly, we must address the accuracy of our models and how they may be affected by the lack of data. The cost model's accuracy is affected greatly by the data included in generating the model. These may very well not represent an average DoD program at all due to the limitations already mentioned in the data. Additionally, due to the nature of the analysis, this model gives the median program estimate. This is most useful when analyzing a portfolio of programs from the directorate level, rather than a single program in an isolated manner. This will allow the decision maker the greatest level of confidence and flexibility when analyzing the results and allocating the resources. Concerning the logistic regression models, we lacked the ability to validate the model by setting aside 20% of the data due to the small sample size. This is an issue when discussing the potential accuracy because we simply could not test how accurate these models may be. As mentioned in Chapter V, we use the bootstrapping method to give a 90% confidence interval for these models to remedy this issue.

Recommendations for Future Research

There are three areas that we believe this research can be expanded upon and ultimately improved. Firstly, as identified in prior research and already mentioned by this research, the use of IOC as a termination date for regression analysis inherently introduces a level of known variability that is unaccounted for in our models. By accounting for this known variability, future researcher could increase the fidelity on actual major drivers of cost and schedule. This new metric for IOC would potentially look like a certain percentage of units delivered, which could standardize all programs in the dataset. Standardizing this termination point could not only influence how researchers analyze future DoD programs, but also potentially influence how IOC is defined in a program.

The second area for future research we identify is expanding upon our logistic regression models. We use MS B as a proxy for the current APB and define the two thresholds as significant and critical from this estimate, using IOC as the termination point. It could be more useful to include the original APB and apply the significant and critical threshold overruns from this point in the program. This is because programs are simultaneously held accountable by the Nunn-McCurdy Act thresholds of their current and original baselines and can breach either one of them.

The third area for future research we identify is to expand upon the analysis of acquisition reform to include more years in which there were major reforms. This would expand upon the 1985 or Later for MS B Start variable identified by Brown et al. (2015) and identify if the programs started in the different eras of acquisition reform are significantly different from one another. This could potentially identify whether or not acquisition reform has helped the acquisition process in terms of preventing schedule slippage or cost overruns.

As always, the limitations of a small sample size can be remedied by future research by simply being able to add more programs to the analysis. This may give more insight into true predictive variables for all of the models in this research. It could also identify variables we found predictive that are not predictive to the population as a whole but, rather, an artifact of the data we analyzed.

Chapter Summary

In this chapter, we attempt to wrap up our entire research endeavor by discussing the relevant findings, irrelevant findings, our limitations, and how future researchers can build upon what we have done. Through this we are able to draw an end to what we have accomplished and simultaneously provide a stepping off point for others to further our efforts. The findings contained within this research have the potential to impact future cost analysts and program managers when faced with the dilemma to estimate their portfolio's cost and allocate resources accordingly. To this end, we have provided the beginnings of what can become a thorough breakdown of indicators of cost and schedule growth above and beyond what can be considered acceptable.

We recommend the use of our logistic models as a tool to manage a portfolio of programs in order to gain potential elusive insight into the behavioral characteristics of programs. Additionally, we recommend the use of our cost regression model to analytically estimate the median cost of a program, or portfolio of programs, to use as a cross-check for the will-cost estimate and allocate resources accordingly. Additionally, by using this cost model the cost estimator and program manager can work together to identify potential cost savings to satisfy the should-cost mandate from the Secretary of the Air Force.

Appendix A: Implementation of Will-Cost and Should Cost Management

DEPARTMENT OF THE AIR FORCE
WASHINGTON DC

JUN 15 2011

MEMORANDUM FOR SEE DISTRIBUTION

SUBJECT: Implementation of Will-Cost and Should-Cost Management

In order to gain greater efficiency and productivity in Defense spending, the Under Secretary of Defense for Acquisition, Technology & Logistics (USD(AT&L)) has directed the Military Departments and Directors of Defense Agencies to implement Will-Cost and Should-Cost management for all Acquisition Category (ACAT) I, II, and III programs. Dr. Carter, USD (AT&L), is challenging program managers to drive productivity improvements into their programs during contract negotiation and program execution by conducting Should-Cost analysis. This analysis goes beyond the Federal Acquisition Regulation/Defense Federal Acquisition Regulation Supplement (FAR/DFARS) Should-Cost reviews. FAR/DFARS Should-Cost reviews set realistic objectives for negotiating the immediate contract. The Should-Cost estimate as defined in this implementation memorandum is much broader in definition, covering all government and contract program costs throughout the entire life-cycle. SAF/AQ and SAF/FM fully support the implementation of Will-Cost and Should-Cost management and expect the Air Force acquisition community to embrace the concepts and adjust our management processes immediately.

The Department will continue to set program budget baselines using non-advocate Will-Cost estimates. Air Force guidance and instruction (e.g., AFPD 65-5 and AFI 65-508) describe specific requirements for non-advocate Will-Cost estimates or Service Cost Positions in support of ACAT I milestone decisions. However, the same level of rigor and attention is currently not required for ACAT II and III programs even though they account for about 48 percent of the Air Force acquisition budget. To ensure we exercise the same discipline for these programs that we do for our ACAT I programs, all ACAT II and III programs identified on the Acquisition Master List will present Will-Cost estimates at milestone decisions that have been approved by the appropriate product or logistics center financial management cost estimating organization (FMC). As with ACAT I programs, the non-advocate Will-Cost estimate will be used as the basis for all budgeting and programming decisions. All metrics and reporting external to the department will be based on the Will-Cost estimate.

Program managers must begin to drive leanness into their programs by establishing Should-Cost estimates at major milestone decisions. The Should-Cost estimate is an internal management tool for incentivizing performance to target, and is, therefore, not to be used for budgeting, programming, or reporting outside the department. Therefore, Should-Cost estimate documentation must be marked and treated as For Official Use Only. We recognize program managers have concerns about providing estimates that are lower than the budget, since DoD culture tends to use programming and budgeting to incentivize achievement. That is not the intent of this initiative. Will-Cost estimates are the official program position for budgeting, programming, and reporting.

Program managers are responsible for developing Should-Cost estimates. They should ensure cross-functional involvement in the development of the Should-Cost estimate and they can seek assistance from outside organizations (e.g., the Air Force Cost Analysis Agency or Defense Contract

Management Agency) throughout the development process. This effort does not necessarily require large teams to perform detailed bottoms-up assessments on every ACAT I, II, and III program. In some cases, this level of detailed analysis is extremely beneficial and desired, but we expect Program Executive Officers (PEOs), Designated Acquisition Officials (DAOs), and program managers to consider resources required versus potential benefits to determine the best approach. At a minimum, program managers are expected to identify specific discrete measurable items or initiatives that achieve savings against the Will-Cost estimate.

In accordance with USD AT&L direction, program managers for ACAT I, II and III programs identified on the Acquisition Master List will present Should-Cost estimates at their next major milestone. The Milestone Decision Authority (MDA) will approve all Should-Cost estimates and will expect program managers to manage, report, and track to these estimates. We will provide an annual report to OUSD (AT&L)/ARA on our progress. By 1 Jul 2011, PEOs/DAOs will submit a prioritized plan and timeline for completing Should-Cost estimates on all their ACAT I, II, and III programs not scheduled for a major milestone review in 2011. We recognize a waiver for some of these requirements may make sense. USD(AT&L) will consider and approve waivers for ACAT ID and IAM programs. SAF/AQ and SAF/FM will consider and approve waivers for all ACAT IC/IAC programs. The PEOs/DAOs and product/logistic center FM leads will approve waivers for ACAT II and III programs.

The following Air Force programs have been designated as pilots: JSF (F-35), Global Hawk Blocks 30 & 40, Evolved Expendable Launch Vehicle (EELV), Space Based Infrared System (SBIRS), and Advanced Extremely High Frequency (AEHF) Satellite System. These programs will be the first to actually have funds distributed based on Should-Cost execution baselines. The difference between the funds distributed and the program budget baseline will be held at the Service level. SAF/AQ and SAF/FM will jointly be the decision authority for release of these funds. We will need to capture lessons learned from each of these programs and share them with OSD and the other Services.

The attachment provides additional guidance and clarifies terms, procedures, and reporting requirements associated with this initiative. The guidance will be updated and codified in policy as USD(AT&L) and the Services/Components gain experience with Will-Cost and Should-Cost management. The POCs for this issue are Ms. Ranae Woods, AFCAA/TD, 703-604-0400, ranae.woods@us.af.mil and Mr. Bob Martin, SAF/AECO, 703-588-7177, robert.martin@pentagon.af.mil.

Jamie M. Morin
Assistant Secretary of the Air Force
(Financial Management and Comptroller)

David M. Van Buren
Air Force Service Acquisition Executive

Appendix B: Predictor Variables Investigated in Thesis

- *MS A to MS B Duration (Months) – Continuous Variable*
  - This variable indicates the total time it took in months for a program to complete MS A to MS B according to the last SAR date. In this variable we are only concerned with actual schedule duration data available to the cost estimator at the time of Milestone B/EMD start.

- *Quantity Expected at MS B – Continuous Variable*
  - This variable indicates the estimate of total quantity of weapons systems that were expected to be produced at MS B at the time of the last SAR date.

- *RDT&E $ (M) at MS B Start (BY17) – Continuous Variable*
  - This variable is based on raw total RDT&E dollars (in millions) that were allocated to the program before, and up to the start of, MS B. The dollars were all standardized into the base year, when the research began (BY17).

- *(Projected) % of RDT&E Funding at MS B Start (BY17) – Continuous Variable*
  - This variable is based on the percent of available RDT&E dollars allocated to the program before, and up to the start of, MS B. While this variable is based on a percentage, the dollars that this percentage was derived from were all standardized into the base year, when the research began (BY17).

- *(Projected) Total Program Acquisition Cost (BY17) – Continuous Variable*
  - This variable is the total projected acquisition costs, from MS B to IOC, estimated at MS B or the earliest available program SAR. It serves to identify how large a program is projected to be in terms of cost.

- *Modification – Binary Variable*
  - This variable is concerned with programs whose existence serves as a modification to a pre-existing weapons system. If a weapons system is a modification, it does not necessarily mean it will not have pre-MS B data associated with it. Every program is different and, therefore, it cannot be assumed that a modification will automatically start at MS B.

- *Prototype – Binary Variable*
  - This variable is concerned with programs that create a prototype, or prototypes, of a weapons system before production of that weapons system begins. More than one type of prototype for a weapons system can be created in a given program.

- *Concurrency Planned – Binary Variable*
  - This variable addresses planned concurrency in a given program prior to MS B. Concurrency is the proportion of RDT&E dollars that are authorized

during the same years that Procurement appropriations are authorized. The planned level of concurrency forces managers to make decisions that can lead to [schedule] growth if either too much or too little concurrency is accepted for a given program (Birchler, Christle, & Groo, 2011, p. 246).

- *1985 or Later for MS B Start – Binary Variable*
  - This variable accounts for a time series trend of programs that started their MS B in 1985 or later. It is shown that programs which began development during 1985 or later (considered "contemporary") expend a greater percentage of obligations by their schedule midpoint than the earlier pre-1985 programs. We attribute this difference to the President's Blue Ribbon Commission on Defense (commonly called the Packard Commission) and the subsequent acquisition reforms.

- *Air Force – Binary Variable*
  - This variable identifies if the lead service on the program was the United States Air Force.

- *Navy – Binary Variable*
  - This variable identifies if the lead service on the program was the United States Navy.

- *Army – Binary Variable*
  - This variable identifies if the lead service on the program was the United States Army.

- *Marine Corps – Binary Variable*
  - This variable identifies if the lead service on the program was the United States Marine Corps.

- *Aircraft – Binary Variable*
  - This variable identifies if the weapons system program is an aircraft program, regardless of service it is associated with. The criterion to qualify as an aircraft for this variable is any weapons system whose primary function is flight; both rotary-wing and fixed-wing programs.

- *Fighter Program – Binary Variable*
  - This variable identifies if the weapons system program is a fighter program, or close variation thereof, regardless of service it is associated with.

- *Bomber Program – Binary Variable*
  - This variable identifies if the weapons system program is a bomber program, or close variation thereof, regardless of service it is associated with.

- *Helo Program – Binary Variable*
  - This variable identifies if the weapons system program is a helicopter program, or close variation thereof, regardless of service it is associated with.

- *Cargo Plane Program – Binary Variable*
  - This variable identifies if the weapons system program is a cargo plane program, or close variation thereof, regardless of service it is associated with.

- *Tanker Program – Binary Variable*
  - This variable identifies if the weapons system program is a tanker plane program, or close variation thereof, regardless of service it is associated with.

- *Electronic Warfare Program – Binary Variable*
  - This variable identifies if the weapons system program is an electronic warfare program, or close variation thereof, regardless of service it is associated with. An electronic warfare program, as not to be confused with an electronic system program, differs greatly in its main function(s). A description from Lockheed Martin makes the distinction that it involves the ability to use the electromagnetic spectrum – signals such as radio, infrared or radar – to sense, protect, and communicate. At the same time, it can be used to deny adversaries the ability to either disrupt or use these signals (Electronic Warfare).

- *Trainer Plane Program – Binary Variable*
  - This variable identifies if the weapons system program is a trainer plane program, or close variation thereof, regardless of service it is associated with.

- *Missile Program – Binary Variable*
  - This variable identifies if the weapons system program is a missile program, or close variation thereof, regardless of service it is associated with.

- *Electronic System Program – Binary Variable*
  - This variable identifies if the weapons system program is an electronic system program, or close variation thereof, regardless of service it is associated with. This differs greatly from the previously described electronic warfare variable in that electronic systems programs are principally concerned with the electronic user interface of a system, avionics controls, or other similar applications that primarily support the electronic usability of a system, or system of systems.

- *Submarine Program – Binary Variable*
  - This variable identifies if the weapons system program is a submarine

program, or close variation thereof, regardless of service it is associated with.

- *Ship Program – Binary Variable*
  - This variable identifies if the weapons system program is a surface ship program, or close variation thereof, regardless of service it is associated with.

- *Satellite Program – Binary Variable*
  - This variable identifies if the weapons system program is a satellite program, or close variation thereof, regardless of service it is associated with.

- *ACAT I – Binary Variable*
  - This variable makes the distinction if the program is an ACAT I program or not. This is significant in that ACAT I programs deal with a much larger dollar amount and thus are more susceptible to cost and schedule growth by way of their large-scale and complexity efforts.

- *(Projected) MS C to IOC Duration (Months) – Continuous Variable*
  - This variable indicates the total estimated time, in months, for a program to meet IOC from MS C according to the earliest available SAR estimate. This variable has been found to be predictive of cost growth in the programs studied by Foreman (2007). With this variable, we are concerned with giving the cost estimator the ability to enter in the projected duration, in months, of the gap between MS C and IOC to predict program cost.

- *(Projected) MS C Slip – Binary Variable*
  - This variable indicates whether the program projected date for meeting IOC extends past the initial estimate. Foreman (2007) has found that a slip in MS C is indicative of program cost growth in past research.

- *No MS A Date – Binary Variable*
  - This variable identifies whether a program did not contain a MS A date in the schedule portion of the SAR, but did include funding at least one year prior to MS B. This is used to identify these programs and test that they are not statistically different from the other programs and is not used in a predictive capacity.

- *Small Program – Binary Variable*
  - This variable identifies whether a program's projected total acquisition costs (RDT&E and Procurement) are below $3000 M. This value is determined from analyzing the histogram of the (projected) total program acquisition costs of the programs in our study and coincides closely with the 25% value.

- *Medium Program – Binary Variable*
    - This variable identifies whether a program's projected total acquisition costs (RDT&E and Procurement) are above $3000 M but below $7000 M. This value is determined from analyzing the histogram of the (projected) total program acquisition costs of the programs in our study and coincides closely with the 25% to 50% range.

- *Large Program – Binary Variable*
    - This variable identifies whether a program's projected total acquisition costs (RDT&E and Procurement) are above $7000 M but below $17500 M. This value is determined from analyzing the histogram of the (projected) total program acquisition costs of the programs in our study and coincides closely with the 50% to 75% range.

- *Extra Large Program – Binary Variable*
    - This variable identifies whether a program's projected total acquisition costs (RDT&E and Procurement) are above $17500 M. This value is determined from analyzing the histogram of the (projected) total program acquisition costs of the programs in our study and coincides with the 75% value.

- *(Projected) % Complete at MS B Start – Continuous Variable*
    - This variable is inspired by the % RDT&E variable and serves to project the percent that a program is complete, to IOC, when MS B occurs. It is calculated by dividing the projected duration from MS B to IOC by the sum of duration from MS A to IOC and projected duration from MS B to IOC. This serves to indicate where the program managers believe the program is in terms of schedule completeness. It could indicate program maturity level.

Appendix C: E-mail Correspondence Concerning Use of Jimenez's Predictive Model

File | Message

Ignore | Delete | Reply | Reply All | Forward | Meeting | Move to: ? | To Manager
Junk | More | Team E-mail | Done | Reply & Delete | Create New

Delete | Respond | Quick Steps

Follow up. Start by Friday, August 19, 2016. Due by Friday, August 19, 2016.
You forwarded this message on 8/19/2016 9:10 AM.

Red Category

From: BROWN, GREGORY E Capt USAF AFMC AFLCMC/WISF <gregory.brown.34@us.af.mil>
To: Trudelle, Ryan C 1st Lt USAF AETC AFIT/ENV
Cc: JIMENEZ, CHRISTOPHER A Capt USAF HAF AF COST ANA AG FO/FMC
Subject: RE: Schedule Estimate for KC-46
Signed By: gregory.brown.34@us.af.mil

Message | KC-46 Schedule Model Test.xlsx

Ryan,

Concerning the KC-46A, the model predicts 80.4 months (Nov 2017), and at the time that Chris graduated the KC-46A was going to take 78.8 months (Sep 2017). It was crazy how close they two dates were (see attached). However, I believe the KC-46A has since further slipped, with the new IOC being Jan 2018 (~84 months).

The more recent program that Chris' model was used for was the JSTARS Recap. I'm not sure whether the program office actually employed Chris' model, I'm thinking it was probably AFCAA that used it when conducting their independent cost estimate for the Recap. Chris (cc'd) should have more information. And if turns out that the program used it as well, I'll try to dig up a name for one of the cost people that works that program.

Hope this helps.
Greg

Bibliography

Birchler, D., Christle, G., & Groo, E. (2011). Cost implications of design/build concurrency. *Defense Acquisition Research Journal*, 18(3), 237–246.

Bolten, J.G., Leonard, R.S., Arena, M.V., Younossi, O. & Sollinger, J.M. (2008). *Sources of Weapon System Cost Growth: Analysis of 35 Major Defense Acquisition Programs (MG 670)*. Santa Monica, CA: RAND Corporation.

Brown, G. E., White, E. D., Ritschel, J. D., & Seibel, M. J. (2015). Time Phasing Aircraft R&D Using the Weibull and Beta Distributions. *Journal of Cost Analysis and Parametrics*, 8(3), 150–164. http://doi.org/10.1080/1941658X.2015.1096219

Burgess, E., & Krause, C. (2014, August). *Integrated estimating relationships*. Paper presented at the NASA 2014 Cost Symposium, Hampton, VA. Retrieved from https://www.nasa.gov/sites/default/files/files/09_Burgess_NASA_Cost_Symposium_2014_IERs_Final_Tagged.pdf

Carroll, R. J., & D. Rupert (1981). On prediction and the power transformation family. *Biometrika*, 68(3), 609-615. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.536.2459&rep=rep1&type=pdf

Carter, A. B., & Mueller, J. (2011). Should Cost Management: Why? How? *Defense AT&L*, 40(5), 14–18. http://www.dau.mil/pubscats/ATL%20Docs/Sep Oct11/Carter_Mueller.pdf

Cook, D. (1977). Detection of Influential Observation in Linear Regression. Technometrics, 19(1), 15–18.

Defense Acquisition University (DAU) "IOC Definition" (2016). Retrieved July 12, 2016, from https://dap.dau.mil/glossary/pages/2048.aspx

Deitz, D., Eveleigh, T. J., Holzer, T. H., & Sarkani, S. (2013). Improving program success through systems engineering tools in the pre-milestone B acquisition phase. *Defense Acquisition Research Journal*, 20(3), 283–308.

Lockheed-Martin "Electronic Warfare Definition" (2016). Retrieved September 6, 2016, from http://www.lockheedmartin.com/us/products/electronic-warfare.html

Foreman, J. D. (2007). *Predicting the effect of longitudinal variables on cost and schedule performance* (Master's thesis). Defense Technical Information Center (DTIC) (ADA463520), Fort Belvoir, VA. http://www.dtic.mil/get-tr-doc/pdf?AD= ADA463520

Gailey III, C. K. *Predictive Power for Program Success from Engineering and Manufacturing Development Performance Trends*. Fort Belvoir VA: Defense Acquisition University, November 2002 (TR 1-02).

Gaudard, M., Ramsey, P., & Stephens, M. (2006). *Interactive data mining and design of experiments: The JMP® partition and custom design platforms*. Cary, North Carolina: North Haven Group, LLC http://islab.soe.uoguelph.ca/sareibi/PROJECTS_dr/GRAD_FUTURE_dr/docs/Interactive_DataMining.pdf

Genest, D., & White, E. (2005). Predicting RDT&E Cost Growth. *The Journal of Cost Analysis & Management*, 7(1), 1–12.

Harmon, B. R. (2012, September). *The limits of competition in defense acquisition*. Paper presented at the Defense Acquisition University Research Symposium, Fort Belvoir, VA. http://www.dau.mil/Research/symposiumdocs/Harmon-Near substitutes paper.pdf

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70. https://www.ime.usp.br/~abe/lista/pdf4R8xPVzCnX.pdf

Hough, P. (1992). *Pitfalls in Calculating Cost Growth from Selected Acquisition Reports*. Santa Monica CA: RAND, 1992 (N-3136-AF)

James, D. (2015). *Reinventing the Aerospace Nation*. In Air and Space Conference and Technology Exposition. Retrieved from http://www.af.mil/AboutUs/AirForceSeniorLeaders/SECAF.aspx

Jimenez, C. A., White, E. D., Brown, G. E., Ritschel, J. D., Lucas, B. M., & Seibel, M. J. (2016). Using Pre-Milestone B Data to Predict Schedule Duration for Defense Acquisition Programs. *Journal of Cost Analysis and Parametrics*, 9(2), 112–126. http://doi.org/10.1080/1941658X.2016.1201024

Kozlak, S. (2016). *Predicting Cost Growth Using Program Reviews and Milestones for DOD Aircraft*. (Unpublished Master's Thesis)

Lucas, B. M. (2004). *Creating Cost Growth Models for the Engineering and Manufacturing Development Phase of Acquisition Using Logistic and Multiple Regression*. (Master's thesis). Defense Technical Information Center (DTIC) (ADA422915), Fort Belvoir, VA. http://www.dtic.mil/get-tr-doc/pdf?AD= ADA422915

McPherson, R. A., & Pincus, M. R. (2016). *Henry's clinical diagnosis and management by laboratory methods* (23rd ed.). Amsterdam, Netherlands: Elsevier.

Monaco, J., & White, E. (2005). Extending Cost Growth Estimation to Predict Schedule Risk, *The Journal of Cost Analysis and Management*, 1:13, 2006.

Monaco, J., & White, E. (2005). Investigating Schedule Slippage, *Defense Acquisition Review Journal*, 12(2), 177–193, 2005.

Neter, J., Kutner, M. H., Nachtseim, C. J., & Wasserman, W. (1996). *Applied linear statistical models* (4th ed.). Boston, MA: McGraw-Hill Companies.

Packard, D. (1986). President's Blue Ribbon Commission on Defense Management.

Reig, R. (1995). *A Decade of Success and Failures in the DoD Acquisition System*. Program Manager, 1, 27–29.

Schwartz, M., & O'Connor, C. V. (2016). *The Nunn-McCurdy act: Background, analysis, and issues for Congress* (Report No. R41293). Washington, DC.: Congressional Research Service.
Retrieved from https://fas.org/sgp/crs/natsec/R41293.pdf

Stine, R. (1995). *Graphical Interpretation of Variance Inflation Factors*. The American Statistician, 49(1), 53–56.

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19(3), 227–229.

Tisdel, J. E. (2006). *Small sample confidence intervals in log space back-transformed from normal space*. (Master's thesis). Defense Technical Information Center (DTIC) (ADA450276), Fort Belvoir, VA. Thesis. Air Force Institute of Technology, 2006. http://www.dtic.mil/get-tr-doc/pdf?AD=ADA450276

Tyson, K. W., Harmon, B. R., & Utech, D. M. *Understanding Cost and Schedule Growth in Acquisition Programs.* Alexandria VA: Institute for Defense Analysis Paper 2967, July 1994 (AD-A284321).

Tyson, K. W., Nelson, J.R, & Om, N. I., Palmer, P. R. *Acquiring Major Systems: Cost and Schedule Trends and Acquisition Initiative Effectiveness.* Alexandria VA: Institute for Defense Analysis Paper 2201, March 1989 (AD-A212538).

Weisstein, E. W. (2015). Bonferroni Correction. Wolfram Research, Inc. Retrieved from http://mathworld.wolfram.com/BonferroniCorrection.html

White, E. D., Sipple, V. P., & Greiner, M. A. (2004). Using logistic and multiple regression to estimate engineering cost risk, *Journal of Cost Analysis and Management*, 6(1), 67–79. http://www.tandfonline.com/doi/abs/10.1080/15411656.2004.10462248

Vita

Captain Ryan C. Trudelle graduated with a Bachelor of Science (B.S.) degree from the United States Air Force Academy in the Class of 2012. Following the completion of his undergraduate degree, he was commissioned as an Officer in the United States Air Force.

During his Air Force career, Captain Trudelle has gained a variety of experience across the spectrum of Air Force Acquisition, to include budget and cost analysis. Upon graduation from the Air Force Institute of Technology with a Master of Science (M.S.) degree in Cost Analysis, he will be assigned as a Cost Analysis Officer at the Space and Missiles Center in Los Angeles, CA.

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

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
23-03-2017	Master's Thesis	Sep 2015 – March 2017

TITLE AND SUBTITLE	
Using Multiple and Logistic Regression to Estimate the Median Will-Cost and Probability of Cost and Schedule Overrun for Program Managers	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Trudelle, Ryan C., Captain, 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 641 WPAFB OH 45433-8865	AFIT-ENC-MS-17-M-231
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
Air Force Life Cycle Management Center 1865 4th Street, Wright-Patterson AFB, OH 45433 Phone: (937) 656-8474; E-mail: gregory.brown.34@us.af.mil ATTN: Capt Gregory Brown	AFLCMC/WISF **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
The main concern of a program manager is to manage the cost, schedule, and performance triad of a program. Historically, programs tend to meet the performance aspect at the expense of cost or schedule, or both. This research gives the acquisition community a set of tools that enables them to impartially analyze the cost and schedule of their programs, helping to mitigate these issues. Five regression models encompass this toolset; one to estimate the median program cost and four to identify the probability of realizing a given overrun. The cost model explains 81% of the variation in program acquisition using seven predictor variables available to the estimator at the time of MS B start. Four logistic models estimate the probabilities that a program may identify as a program that experiences cost and schedule overruns of specific magnitudes from their MS B estimate. These models predict the group the program may reside in with an accuracy of at least 0.79 probability and use multiple predictor variables available at MS B. With these tools the program manager has the ability to preemptively identify potential problems in their program based on the program's characteristics, potentially saving millions in cost and schedule overruns.

15. SUBJECT TERMS
Cost Analysis, Statistics, Regression Analysis, Defense Acquisitions, Schedule, Finance, Program Management

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			White, Edward D., Ph.D., AFIT/ENC
U	U	U	UU	121	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636, ext 4540 (NOT DSN) (edward.white@afit.edu)

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