Predicting Schedule Duration for Defense Acquisition Programs:

Program Initiation to Initial Operational Capability

THESIS

Christopher A. Jimenez, B.S.

Captain, USAF

AFIT-ENC-MS-16-M-161

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

AFIT-ENC-MS-16-M-161

Predicting Schedule Duration for Defense Acquisition Programs: Program Initiation to Initial Operational Capability

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

Christopher A. Jimenez, B.S.

Captain, USAF

March 2016

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENC-MS-16-M-161

Predicting Schedule Duration for Defense Acquisition Programs:

Program Initiation to Initial Operational Capability

Christopher A. Jimenez, B.S.

Captain, USAF

Committee Membership:

Dr. Edward D. White, Ph.D
Chair

Lt Col Jonathan D. Ritschel, Ph.D
Member

Lt Col Brandon M. Lucas, Ph.D
Member

Capt Gregory E. Brown, USAF
Member

AFIT-ENC-MS-16-M-161

Abstract

Accurately predicting the most realistic schedule for a defense acquisition program is an extremely difficult challenge considering the inherent risk and uncertainties present in the early stages of a program. To minimize the risk of underestimating or overestimating a program's schedule, the program manager requires a transparent, unbiased method of schedule estimation. Through the application of multiple regression modeling, we provide the program manager with a statistical model which predicts schedule duration from Program Initiation (Milestone B) to the Initial Operational Capability of the program's deliverable system. Our model explains 42.9 percent of the variation in schedule duration across the historical data from a sample of 56 defense programs from all military services. Statistically significant predictor variables include whether a program is a new effort or modification to an existing program, the year of Milestone B start as it relates to changes in defense acquisition reform policy, and the amount of raw funding (adjusted for inflation) prior to Milestone B for a program. Our strongest predictor variable, percent of total RDT&E funding occurring prior to Milestone B, indicates that increased funding for pre-Milestone B technology risk reduction may shorten a program's schedule duration to Initial Operational Capability.

Acknowledgments

In being a first generation college graduate, I would like to start by thanking God for all of the opportunities I have been able to come across up to this point in my life. Thank you to my mother and father for their inspiration and encouragement along the way. Also, a special thank you to my Godmother, extended family, mentors, and friends as they were a big part of my greater support system that kept me going on a daily basis. To my little brother, thank you for keeping up to date every so often with what was going on in the "real world".

I would also like to warmly thank my research advisor, Dr. Edward White, for his guidance and support during my thesis work. His profound knowledge of statistical model building and regression analysis resulted in countless improvements to this research. Additionally, I wish to express the deepest gratitude towards Lt Col Jonathan Ritschel, Lt Col Brandon Lucas, Capt Gregory Brown, and Mike Seibel; their combined experience added much value in ensuring this research is applicable to the greater cost analysis community.

Christopher A. Jimenez

Table of Contents

Page

Abstract iv
Acknowledgments v
Table of Contents vi
List of Figures ix
List of Tables xi

I. Introduction 1

General Issue 1

Specific Issue 3

Scope and Limitations of Research 3

Research Objectives 5

Research Questions 6

Summary 6

II. Literature Review 8

Chapter Overview 8

Research Findings 8

Chapter Summary 30

III. Methodology 32

Chapter Overview 32

Database 32

Selected Acquisition Report (SAR) Data 36

Response Variable 36

Search for Predictors of Schedule Duration 37

Predictor Variables 38

Validation Pool 42

Exploratory Data Analysis 42

Multiple Regression Analysis.....46

Validation of Multiple Regression Model 48

Chapter Summary 49

IV. Results and Analysis 50

Chapter Overview 50

Preliminary Multiple Regression Model 50

Validating Model Assumptions 52

Analysis of Predictor Variables 57

Performance of Multiple Regression Model 59

Validation of Multiple Regression Model 60

Chapter Summary 66

V. Conclusions and Recommendations 67

Chapter Overview 67

Research Questions Answered 67

Findings 69

Limitations 69

Recommendations for Future Research 71

Chapter Summary 72

Appendix A: List of Acronyms 74

Appendix B: Implementation of Will-Cost and Should-Cost Management 76

Appendix C: List of 56 Programs in Research Database.....78
Appendix D: Contractors in Research Database.....80
Appendix E: All Data for 56 Programs in Research Database 81
Appendix F: Response and Predictor Variables 82
Appendix G: Validation Pool of 11 Programs.....85
Appendix H: Preliminary/Main Model with 45 Programs 86
Appendix I: Validated/Final Model with 56 Programs.....87
Bibliography 89
Vita….....94

List of Figures

Page

Figure 1: Defense Acquisition Program Schedule with MS-B to IOC Depiction 4

Figure 2: Frequency of Reasons for Schedule Variance by Category based on CPR's ... 18

Figure 3: Average Schedule Variance (work days) by Category based on CPR's 18

Figure 4: Drezner and Smith's Sixteen Schedule Factors 25

Figure 5: Harmon, Ward, and Palmer Schedule Drivers for Tactical Aircraft 27

Figure 6: Harmon and Ward Schedule Drivers for Air-Launched Missiles 28

Figure 7: Preliminary Multiple Regression Model Output 51

Figure 8: Preliminary Model VIF Scores 53

Figure 9: Display of Cook's D Plot 54

Figure 10: Studentized Residuals 55

Figure 11: Shapiro-Wilk's Test 56

Figure 12: ANOVA Output 56

Figure 13: Parameter Estimates for Predictor Variables 57

Figure 14: Model R^2 and Adjusted R^2 Values 60

Figure 15: RDT&E $ (M) at MS-B Start (BY16) Quantiles 61

Figure 16: % RDT&E Funding at MS-B Start (BY16) Quantiles 61

Figure 17: MdAPE and MAPE of Final Model 62

Figure 18: MdAPE and MAPE of Validation Pool 62

Figure 19: Bivariate Plot of Model with 45 Programs 63

Figure 20: Bivariate Plot of Validation Pool with 11 Programs 64

Figure 21: Final Model with all 56 Programs 65

Figure 22: Pie Chart of Standard Betas..... 68

List of Tables

Page

Table 1: Cost Factors Associated with Technological Maturity..... 11

Table 2: Programs with High and Low Schedule Growth in Development 22

Table 3: Programs with Low and High Cost Growth in Total Program..... 22

Table 4: Process of Database Filtering 35

Table 5: Breusch-Pagan Test Results 56

Predicting Schedule Duration for Defense Acquisition Programs: Program Initiation to Initial Operational Capability

I. Introduction

General Issue

As of the "Implementation of Will-Cost and Should-Cost Management" policy memorandum in June 2011, the Air Force employs a 'will-cost' analysis and 'should-cost' analysis to all Acquisition Category (ACAT) I, II, and III programs as a way to try to realize cost savings through operational efficiencies found in the defense acquisitions process (Implementation of Will-Cost and Should-Cost Management, Appendix B). However, in the Air Force, as well as the other Department of Defense (DoD) services, no similar analysis for schedule duration has existed as a policy for trying to optimize the timeliness of an acquisition.

On 15 September, 2015 at the annual Air Force Association conference that was held in National Harbor, MD, Secretary of the Air Force (SECAF) Deborah Lee James introduced the Air Force's newest acquisition strategy, an initiative she called 'should-schedule'. "The should-schedule approach will work in a similar manner to an acquisition management tool the service has been using called 'should-cost'. Unlike should-cost, the new should-schedule strategy will focus on delivery time. We asked ourselves, 'Can we develop a structure that challenges us and our industry partners to deliver [weapons systems] faster than the schedule determined as part of the independent cost estimate? If we can collectively beat the historical developmental schedules and reward behavior in

government and industry that speeds things up, we have a real chance to make a difference," Secretary James emphasized (James, 2015).

Secretary James and the should-schedule initiative provide for a heightened focus on schedule duration estimating in the cost analysis community. Many program cost estimates are created under the assumption of a static schedule, which can create extrapolated problems in the program if the estimated schedule of a program gets delayed or derailed. Furthermore, research by the RAND Corporation found that increases in schedule effort tend to be the reason for increases in the cost of acquiring a new weapons system due to, at a minimum, increased inflation and overhead factors (Drezner and Smith, 1990:1).

Accurately establishing the most realistic schedule for a program, especially at the official initiation of a program, is an extremely difficult task considering the inherent risk and uncertainties that are present in the early stages of a program. Programs that decide to use an unnecessarily lengthy schedule as a program strategy run the risk of delaying the level of technological advancement that may be critical to national safety.

However, accelerated program schedules increase the risks of unscheduled delays and expensive rework and retooling costs, especially if a problem is found later in the accelerated program schedule. A recurring theme of defense critics however is that most programs err on the side of being too lengthy and that policy reforms should be introduced to shorten the cycle (Drezner and Smith, 1990: iii). Secretary James' should-schedule initiative, along with the push for a greater focus on program scheduling methodology, may be the kind of policy reform Drezner and Smith were alluding to 25 years ago.

Specific Issue

Past research on schedule is relatively limited at AFIT, mostly because students in the Graduate Cost Analysis (GCA) program traditionally tend to focus their research efforts on predicting and optimizing costs, rather than schedule. Current Air Force practice is for cost estimators to either rely on subject matter expert (SME) opinion to evaluate the schedule risk levels of different program factors, or perform an analogous schedule estimate based on a comparable project that has been previously completed. These methods of "best guess" are the current standard applied to arrive at the estimated schedule of a program.

As it currently stands in the Air Force, there is no quantitatively-focused method used for predicting schedule duration of a program that is driven by the data of past weapons systems. This is the first research to be conducted at AFIT that is focused on predicting a program's actual schedule duration based on historical data and mathematical modeling. Tangentially related, Monaco (2005) looks at identifying if a program runs the risk of schedule delay, and then predicting the amount of schedule delay for that specific program after it has experienced a schedule delay; his research employed the use of a two-step mathematical modeling procedure.

Scope and Limitations of Research

The scope of this research is limited to predicting schedule duration in months for defense acquisition programs from program initiation, which is the start of Engineering and Manufacturing Development (EMD), to Initial Operational Capability (IOC). Official program initiation happens when the EMD phase starts, which is at Milestone B. IOC is the state achieved when a capability is available in its minimum usefully deployable

form. At IOC, capability may be fielded to a limited number of users with plans to extend to all intended users incrementally over a period of time. Declaration of IOC may imply that the capability will be further developed in the future, for example by modifications or upgrades to improve the system's performance, deployment of greater numbers of systems (perhaps of different types), or testing and training that permit wider application of the capability (DAU, 2015).

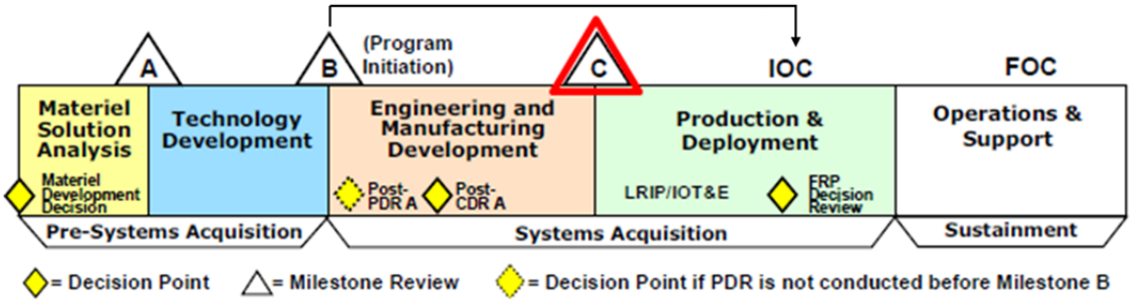


Figure 1: Defense Acquisition Program Schedule with MS-B to IOC Depiction

In our research, we believe that more value could be extracted in seeking to predict schedule duration from MS-B to IOC, instead of seeking to predict schedule duration from MS-B to Production and Deployment (P&D) start which is at Milestone C (MS-C). This is due to the fact that the start of the P&D phase is not always as clearly defined as the start of the MS-B in the acquisition life cycle. Common practice is to have both EMD and P&D run concurrent for some time in the acquisition life cycle, with the intent of having a system being produced while it is simultaneously being developed (Birchler et al., 2011). Because of this practice, decision makers have a less stringent proxy to beginning the P&D phase compared to EMD phase. Thus, if EMD phase is still going on while P&D phase begins, the concurrency between the two phases may present a lack of clarity in the distinction of the two phases to be able to make a sound decision under the program complexities at that point. Furthermore, commanders and decision

makers may be more concerned with the time to IOC of a weapons system, since it directly relates to fielding a capability earliest in support of critical mission needs (DAU, 2015).

Since we seek to predict MS-B to IOC of a program, a limitation is that we look to only include programs that have pre-MS-B data available. We limit our database further to include only unclassified programs that completed the IOC phase of an acquisition. For this data, we use the Selected Acquisition Report (SAR), maintained by the Office of the Secretary of Defense (OSD), which provides reported in-depth finance and schedule data for selected programs (Brown et al., 2015). We also give extra focus on Research Development Test & Evaluation (RDT&E) funding, as it is the funding deployed for both pre-MS-B and MS-B efforts. The detail and availability of the SARs provide the appropriate information needed to build a proprietary database necessary for this research.

Research Objectives

Our major objective is to have the mathematical model developed in this research to be used as a tool in the cost analysis community. The mathematical model employed for this research involves a multiple regression model that provides an output value in months. For the purpose of this study, the output from the multiple regression analysis encompasses overall time duration in months, starting at MS-B, through EMD, P&D, Low-Rate Initial Production (LRIP), and Initial Operational Test and Evaluation (IOT&E) phases, up to IOC.

The objective of the multiple regression model we create is to have it stand as a predictive tool that outputs a schedule duration that decision makers can use as a realistic

schedule benchmark for their programs. A readily available application of said schedule benchmark is for decision makers to try to employ operational efficiencies in a program as to try to deliver a program's capability quicker than what the data-driven benchmark suggests. This creates the kind of structure that Secretary James mentioned; one which can challenge the Air Force and industry partners to deliver [weapons systems] faster than the schedule determined as part of the independent cost estimate (James, 2015).

Research Questions

Our research is focused on addressing two research questions. First, we seek to answer the research question, "Can we accurately predict what the schedule duration of a defense acquisition program should be, from MS-B to IOC, using a mathematical model?" Independent of said mathematical model, we analyze explanatory variables from program data in search of answering the question, "Can we show that some explanatory variables are stronger than others when used for predicting a future program's schedule duration?"

Summary

Predicting the schedule duration from MS-B to IOC for programs can reduce program risks and help ensure intended performance capabilities are realized within a specific program's cost and schedule thresholds. In our research, we identify reasons for schedule variance along with potential predictors of schedule variance by conducting a literature review in Chapter II. The literature review provides the necessary foundation for our data collection and database creation in Chapter III. We then conduct preliminary analysis of the data in order to create the multiple regression analysis model that seeks to

predict a program's schedule to from MS-B to IOC. In Chapter IV, we build, test, and validate the multiple regression analysis, as well as provide a meaningful discussion of the results. Finally, in Chapter V, we provide conclusions to our research, and possible follow-on research.

II. Literature Review

Chapter Overview

Developing a major weapons system is risky and full of uncertainty. Requirements, politics, economics, and the system's technological design are just a few of the uncertainties that create risk in this venture. This can materialize in the form of variance between the planned schedule duration and the actual schedule duration of a program. "Excessive schedules have two significant negative effects: U.S. forces may be left without needed capabilities and longer schedules often mean higher costs" (Tyson et al., 1994:S-1).

To begin addressing our research objectives and questions, we start by looking at research that can give us greater insight into the intricate details associated with predicting a program's schedule duration to IOC. In this chapter, we provide an overview of past research conducted on defense acquisition program schedules, particularly as it relates to helping us identify significant characteristics necessary for our answering our research questions and building of a multiple regression model. For structured continuity, we only provide an overview of research findings on program schedules from within the defense acquisition environment. Based on our literature review, we create a foundation from which to start the methodology for predicting schedule duration to IOC, which we describe in Chapter III.

Research Findings

The time required to create a new weapons system from program initiation to IOC is an important element to understand in the acquisition process. Cost and schedule overruns in major weapons systems are continuing problems that plague the acquisition

environment. The following research studies discuss various direct and indirect findings associated with program schedule inaccuracies and overruns and investigates variables that can help predict schedule duration.

Brown, White, Ritschel, and Seibel (2015)

Brown et. al (2015) investigates the minimal methodology in the literature that is provided for estimating the S-curve's parameter values. Brown, White, and Gallagher (2002) resolve this shortcoming through regression analysis, but their methodology has not been widely adopted by aircraft cost analysts, as it is judged as overly broad and not specific to aircraft. Instead, analysts commonly apply the 60/40 "rule of thumb" to aircraft development, assuming 60 percent expenditures at 50 percent schedule.

Using a sample of 26 DoD aircraft programs, Brown et al. (2015) first tests the accuracy of 60/40, discovering that, as a heuristic, the 60/40 cannot account for differences between new start and upgrade programs. Next, they improve upon prior research by using program characteristics to construct an aircraft-specific methodology for estimating parameters. Finally, they conclude the research by comparing the accuracy of their Rayleigh, Weibull, and Beta S-curve models. The Weibull model explains 82 percent of total variation in expenditures, improving the estimation of annual expenditures by nine percent, on average, over the baseline 60/40 model.

For Brown et al. (2015) in particular, three pieces are relevant to our research. First is the acknowledgement of the 60/40 "rule of thumb" that is applied to aircraft development, assuming 60 percent expenditures at 50 percent schedule. This tells us that if such a concept is applied in the aircraft development community, then perhaps a

similar concept surrounding the percentage of early expenditures in a program could be applied to as a potential predictor of schedule.

Second, they acknowledge that while their methodology utilizes budget and schedule data from the latest SAR available for each aircraft development program, they also acknowledge that this does not account for any cost or schedule growth which exists between the aircraft program's first and latest SAR. The assumption of a static schedule contrasts with the "real world", where cost and schedule estimates are rarely clairvoyant (Brown et al., 2015:60). This further emphasizes the need for our research on schedule duration.

Finally, and most uniquely, Brown et al. (2015) finds a significant variable that is centered on defense acquisition reform policy. They show 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. They hypothesize that this difference is due to the President's Blue Ribbon Commission on Defense (commonly called the Packard Commission) and the subsequent acquisition reforms.

Dietz, Eveleigh, Holzer, and Sarkani (2013)

This study focuses on the pre-MS-B process in a defense acquisition. The researchers state that with 70 percent of a system's life-cycle cost set at pre-MS-B, the most significant cost savings potential is prior to MS-B. Pre-MS-B efforts are usually reduced to meet tight program schedules. This article proposes a new Systems Engineering Concept Tool and Method (SECTM) that uses genetic algorithms to quickly identify optimal solutions. Both are applied to unmanned undersea vehicle design to

show process feasibility. The method increases the number of alternatives assessed, considers technology maturity risk, and incorporates systems engineering cost into the Analysis of Alternatives process. While not validated, the SECTM would enhance the likelihood of success for sufficiently resourced programs (Deitz et al.,2013). In Table 1, we analyze a cost estimating relationship (CER) table the researchers created relating the technical maturity of a program as it crosses into MS-B, and a cost multiplier associated with said maturity.

Table 1: Cost Factors Associated with Technological Maturity

Viewpoint	Very Low	Low	Nominal	High	Very High
Lack of Maturity	Technology proven and widely used throughout industry	Proven through actual use and ready for widespread adoption	Proven on pilot projects and ready to roll-out for production jobs	Ready for pilot use	Still in the laboratory
Lack of Readiness	Mission proven (TRL 9)	Concept qualified (TRL 8)	Concept has been demonstrated (TRL 7)	Proof of concept validated (TRL 5 & 6)	Concept defined (TRL 3)
Obsolescence	(Obsolescence not an issue)	(Obsolescence not an issue)	Technology is the state-of-the-practice; emerging technology could compete in future	Technology is stale; new and better technology is on the horizon in the near-term	Technology is outdated and use should be avoided in new systems; spare parts supply is scarce
Cost Multiplier	0.68	0.82	1.0	1.32	1.75

We look to this study as rudimentary justification to collect pre-MS-B data for the purpose of predicting schedule, as the researchers were able to derive predictive factors for programs using data based on technological maturity in the pre-MS-B phase.

Birchler, Christle, and Groo (2011)

Birchler et al. (2011) acknowledges the idea that developing a weapons system while in production does increase program risk and is sometimes cited as a reason for cost growth. This description is known as concurrency in the defense acquisition

community. The researchers explore the relationship between concurrency and cost growth in large weapons programs (Birchler et al., 2011).

The researchers defined concurrency as the proportion of research, development, and test and evaluation appropriations authorized during the same years in which procurement appropriations are authorized. Their results strongly indicate that concurrency does not necessarily predict cost growth. Using multiple regression techniques, the researchers found no evidence supporting this relationship. To investigate other relationships between cost growth and concurrency, they also used a smooth curving technique. These experiments showed that, although the relationship is not strong, low levels of concurrency can be more problematic than higher levels (Birchler et al., 2011).

The findings associated with concurrency not significantly predicting cost growth gives us motivation to investigate concurrency for our research as it relates to predicting schedule duration. Perhaps a program with a planned level of concurrency could be statistically significant in predicting schedule duration.

Giacomazzi III (2007)

This research presents an empirical model of schedule growth to evaluate the impact of acquisition reform efforts, defense budget changes, unexpected inflation, and major contingency operations (war) on schedule growth of major weapon systems. A fixed-effects panel regression model was utilized to describe the schedule performance (using earned value data) of the major weapon system programs managed by the Army, Air Force, and Navy from 1980 to 2002. This research found that unexpected inflation

results in increased schedule growth. In addition, the 2000 revision of the DoD 5000 series accounted for a reduction in schedule growth (Giacomazzi III, 2007: iv)

Because Giacomazzi (2007) found that unexpected inflation results in increased schedule growth, we seek to mitigate any negative inflationary effects to our future model by standardizing any cost and funding information collected in the data gathering process. We seek to standardize said cost and funding information to the Base Year (BY) that our research is being conducted in, and that is in Base Year 2016 (BY16).

Monaco and White (2005, 2006)

Monaco and White's (2005, 2006) research centered on an AFIT SAR database built by Sipple (2002) and modified by Bielecki (2003), Moore (2003), Genest (2004), Lucas (2004), McDaniel (2004), and Rossetti (2004). Their modified research database consisted of 52 program derived from this SAR database. Towards the end of his thesis, Monaco (2005) noted some limitations.

One such limitation pertained to the predictive model. Monaco needed a complete set of data in order for the statistical models to accurately predict the probability and magnitude of schedule growth within the time frame of the EMD phase of acquisition (defined as the interval between MS-B and MS-C). Monaco (2005) found that approximately 27 percent of programs that otherwise met the researcher's criteria did not have a reported value for one of the four necessary 2 schedule dates, e.g. planned and actual dates for MS-B and MS-C. Of the programs missing the appropriate schedule dates, Planned MS-B, Actual MS-B, Planned MS-C, and Actual MS-C did not have complete data 56, 28, 72, and 56 percent of the time, respectively (Monaco, 2005:106).

In addition, Monaco (2005) observed the following missing schedule dates that showed promise as possible predictor variables: First Unit Equipped (FUE), Preliminary Design Review (PDR), Production Contract Award (PCA), Critical Design Review (CDR), EMD Contract Award, and IOC. Due to the fact that the SARs contained missing schedule information, Monaco could not decompose the interval between MS-B to MS-C in order to create predictive models within smaller time frames. In particular, the FUE schedule date also appeared to be very predictable but only present in 19.4 percent of the programs (Monaco, 2005:106).

This is probably the closest research we have found as analogous to our scope of our research. Whereas Monaco (2005) focused on building models to try to predict the probability and magnitude of schedule growth, we feel value could also be added to a program by predicting statistically significant schedule duration beforehand, in that it could mitigate the probability and magnitude of schedule growth before it even happens.

Gailey III (2002)

Gailey (2002) expands the Reig (1995) study's database from 24 to 46 programs that have completed MS-B and reflect 28 program characteristics (Gailey III, 2002:5). The results of the study stated that there appeared to be no correlation between LRIP quantities and the probability that the schedule will slip (Gailey III, 2002:5). This fact contradicts the results of Reig (1995) that Gailey expanded on, which used a smaller database.

Gailey further concluded that of the 28 program characteristics examined, 16 exhibit scatter too extreme to provide reliable predictive power (Gailey III, 2002:11). Although the remaining 12 program characteristics were not discussed specifically, the

findings relevant to this study reiterate that the use of competition and contract type differentiate between a successful and unsuccessful program. Contrary to the previous study, no differences were noted in MS-B success attributable to whether MS-B is completed, which particular contractors are the lead, or whether the program is Joint-Service (Gailey III, 2002:9).

Unger, Gallagher, and White (2001)

Unger et. al (2001) first recommends that the Weibull distribution is a better predictor of RDT&E expenditure profiles than the Rayleigh distribution. Unger tests the ability of both the Rayleigh and Weibull to predict variation and cost and schedule growth, finding that the Weibull outperforms the Rayleigh when fit to individual programs (Unger, 2001:5). The shape of the Weibull suggests a more front-loaded profile. However, in his findings, Unger annotates a significant limitation of his model: no method currently exists to estimate the Rayleigh and Weibull parameters for future programs. Both this study and the work by Brown et al. (2015) share the common idea that front-loaded funding for a program generally relates to lowering schedule growth.

Joint Strike Fighter (JSF) (2000)

Pioneered by the National Aeronautics and Space Administration (NASA) and adopted by the Air Force Research Laboratory (AFRL), Technology Readiness Level (TRL) was used to determine the readiness of technologies incorporated into a weapon or other type of system (Rodrigues, 2000:9). Measured on a scale of one to nine, the lower the level of maturity when a technology was included in a development program, the higher the risk that it would cause problems, such as schedule delays in the future (Rodrigues, 2000:8).

According to NASA, AFRL, and others in DoD, a level of seven enables a technology to be included in a development program with acceptable risk (Rodrigues, 2000:9). TRLs were also used in prior work to assess the impact of technological maturity of product outcomes. A review of 23 different technologies into new product and weapon systems designs within DoD and the commercial sector determined that cost and schedule problems raise when programs start with technologies at low readiness levels and it conversely showed that programs met product objectives when the technologies were at higher levels of readiness (NASA, 2002). Perhaps TRL of a program could serve to potentially explain predicted schedule duration at different TRL levels going into MS-B.

Cashman (1995)

Cashman (1995), in his thesis, addresses three objectives: identifying actual reasons for schedule problems across large Air Force system development efforts, quantifying the importance of each category of reasons in terms of frequency and severity in order to determine the categories of reasons most and least deserving of management attention, and demonstrating that the reasons are not program unique but common across system development efforts (Cashman, 1995:34).

Cashman used data available in Cost Performance Reports (CPRs) located within the Aeronautical Systems Center cost library with funding over $40M limited to the EMD phase specifically. The sample consisted of 22 system development efforts that were ongoing or ended after 1984, described by 549 instances of schedule problems from 1982-1994 relating to aircraft/missile, simulator, aircraft equipment, and aircraft upgrade (Cashman, 1995:25 and 35).

In order for meaningful identification of the reasons for schedule problems, and the quantification of those reasons, it is necessary to group data into categories. "As each reason for schedule problems and associated quantitative information was extracted from the CPR, the reason was categorized based on its wording and the researcher's five years of experience as an Air Force project manager" (Cashman, 1995:31).

It is also noted by the researcher that reasons for schedule problems were not program specific but common across most development efforts. "While all 22 development efforts did not experience all 20 categories of reasons for schedule problems, no category appeared on only one effort, and on average, categories appeared on 9.1 efforts" (Cashman, 1995:69).

Also noted by Drezner and Smith's (1990) factors affecting schedules were technical difficulty and concept stability. One reason for continued schedule overrun in the procurement of major weapons systems over the years is the low level of technical maturity of the system when it proceeds into the EMD phase. Once the development phase begins, the government incurs a large fixed investment in the form of human capital, facilities, and materials. Any changes thereafter may negatively affect schedule duration.

In Figures 2 and 3, we see the chart of reasons for schedule variance based on observations, as well as time duration of schedule variance in work days per category. We look to this accumulated information regarding schedule variance as a group of potential independent variables that could prove to be statistically significant in building our multiple regression model that seeks to predict schedule duration.

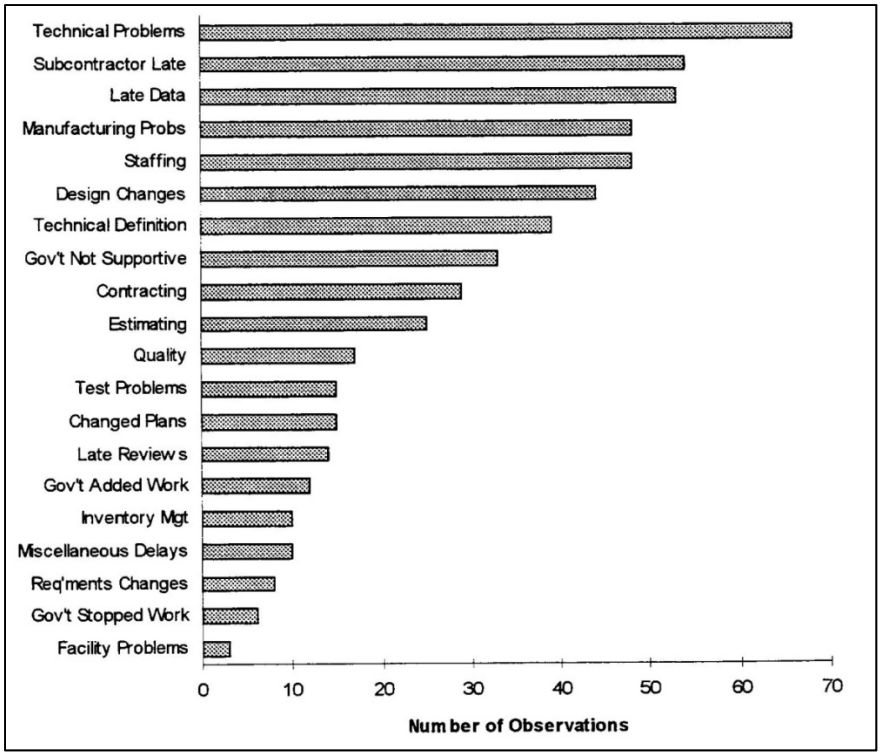


Figure 2: Frequency of Reasons for Schedule Variance by Category based on CPR's (Cashman, 1995:61)

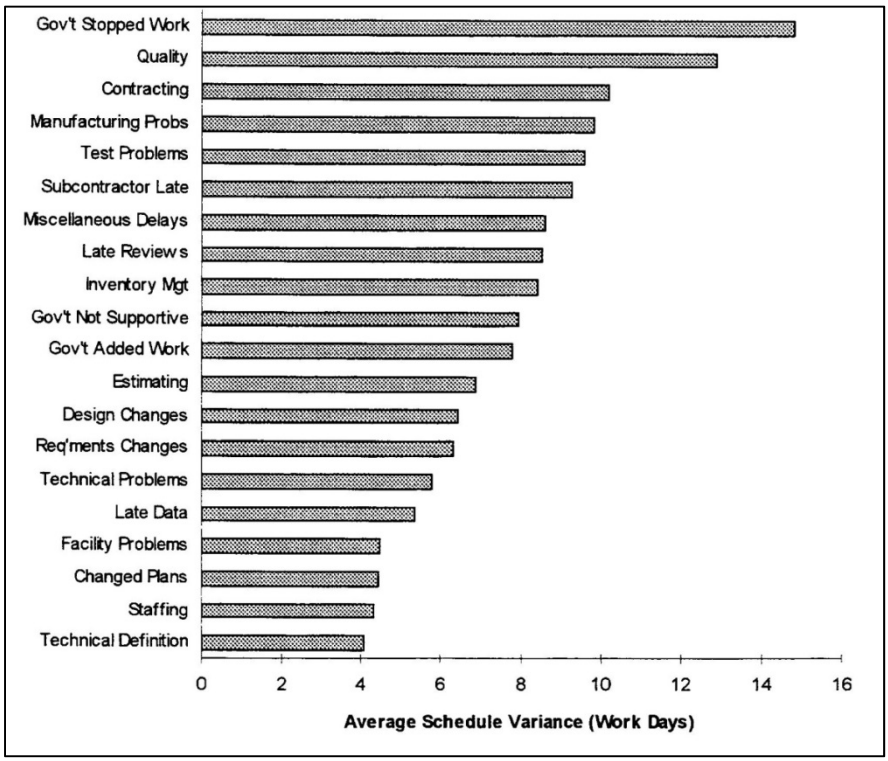


Figure 3: Average Schedule Variance (work days) by Category based on CPR's (Cashman, 1995:6)

Harmon and Om (1995)

This study is conducted by the Institute for Defense Analyses (IDA). The data collected consists of 22 missile programs with substantial developments from the mid-1960s to the 1990s. The breakdown of the 22 programs is: eight surface-launched interceptors, seven air-launched interceptors, and seven air-launched surface-attack missiles (Harmon and Om, 1995: I-2). Although the focus of this study is on interceptor missiles, inclusion of the attack missiles is used because attack missile programs tend to be influenced by the same drivers and the missiles hardware also share many attributes (Harmon and Om, 1995:II-1).

The 22 programs offer a variety of types in both program and missile attributes. Ten of the 22 programs are modification programs based on previously developed missiles (Harmon and Om, 1995:II-9). Development program schedules are decomposed into 4 periods: 1) Time to first guided launch as measured from development start to first guided launch, 2) Length of the development flight test program as measured from the first guided launch to the end of initial operational testing, 3) Early production time as measured from long-lead and full-funding release for the initial production lots to the first production deliveries for those lots, and 4) Program length from first launch as measured by the time from first guided launch to first production delivery (Harmon and Om, 1995:I-3).

The data for this study encounters the same variability in the data and therefore uses delivery date of the first production missile to mark the end of development (Harmon and Om, 1995:I-3-I-4). Although emphasis is placed on both pre-EMD and the EMD phase of the acquisition cycle, schedule intervals in the concept exploration phases

and the demonstration and validation phase prior to EMD are often highly dependent upon political factors and therefore not emphasized. Pre-EMD prototype intervals on the other hand are an exception (Harmon and Om, 1995:II-1).

The researchers originally wanted to develop a single equation to estimate the interval of EMD defined as the period from EMD start to delivery of the first production missile (Harmon and Om, 1995:III-1). "Unfortunately, the determinants of time to first launch and time from first launch to first production are just too different" (Harmon and Om, 1995:III-24). Instead they choose the interval between first guided-launch and the first production delivery (Harmon and Om, 1995:III-1).

According to the research, time to first launch is a function of technological variables whereas time from first launch to first production is a function of the number of missiles launched in flight test, the rate at which they are launched, the overlap between production start and flight test, and production time (Harmon and Om, 1995:III-25). "Our hypothesis was that the terminal guidance system, generally the highest value item and most technologically difficult development item, would pace overall missile development" (Harmon and Om, 1995:II-9). The one program attribute that serves most important in determining length of the development effort is the number of missiles launched during flight tests (Harmon and Om, 1995:II-9).

Tyson, Harmon, and Utech (1994)

Unrelated to the four previous IDA studies, IDA performed an analysis on 20 tactical missile and seven tactical aircraft programs with the objective to describe costs and schedule growth patterns associated with the acquisition of selected major systems, identify reasons for the growth, and develop a way to predict growth in ongoing

development and early production phases (Tyson, et al., 1994:iii). Data used for this study comprises information obtained from SARs, historical memoranda to support DoD program reviews, and from summaries of program data (Tyson, et al., 1994:S-1).

The study finds that programs take from 50 to 137 months from Milestone II to IOC with only two of the twenty tactical missile programs finishing on time and the highest development schedule growth exceeding its plan by 180 percent (Tyson et al., 1994:S-2). Only two programs complete under budget with two other programs exceeding their cost two-fold (Tyson et al., 1994:S-2).

The researchers examine the characteristics of programs with the highest and lowest schedule and cost growth. The results are located in Table 1 and Table 2. (Tyson et al., 1994:S-3 and S-4). The researchers state that keys to preventing schedule growth in development are technical realism and a willingness to make tradeoffs and the keys to preventing overall cost growth are correctly estimating the degree of technical difficulty in the programs and maintaining the planned production schedule (Tyson et al., 1994:S-2). The growth for aircraft is less dispersed than missile programs for multiple reasons. In particular, they note this reason is due to the increased management scrutiny aircraft programs incur and a protection from schedule stretch (Tyson et al., 1994:S-2).

Another finding from this research is that the major determinant of development schedule growth is an increase in quantity; the need to produce more items for testing than planned (Tyson et al., 1994:S-5 and S-6). Contrary to the 1990 RAND study, the researchers in this study find a relationship between cost growth and schedule growth in both development and production (Tyson et al., 1994:S-6; and Drezner and Smith, 1990:45).

Table 2: Programs with High and Low Schedule Growth in Development (Tyson et al., 1994:S-3)

Program	% Growth	Characteristics
Low Growth		
TOW 2	0%	Follow-on system
Sidewinder AIM-9M	1%	Follow-on system to fulfill goals of AIM-9L
		Learned from unrealistic estimate of prior system
MLRS	6%	Urgent program
		Competitive prototype
		Requirements/schedule tradeoff made in favor of schedule
High Growth		
Phoenix AIM-54A	94%	Problems resolved in development, not allowed to spill over into production
		Testing delays
		Delays in aircraft platform
Maverick AGM-65D/G	98%	Funding cut slowed development, allowed technology to catch up
		Prototype
		Vigorous testing program
AMRAAM	129%	Prototype showed infeasibility of approach
		High concurrency, urgent program
		Rushed testing
Sidewinder AIM-9L	148%	Urgent program, with fly-before-buy strategy
		Technical problems, with increased development quantity
Sparrow AIM-7F	180%	Joint service program, with technical disagreements
		Underestimation of technical difficulty (vacuum tube to solid state)
		Vigorous testing program

Table 3: Programs with Low and High Cost Growth in Total Program (Tyson et al., 1994:S-4)

Program	% Growth	Characteristics
Low Growth		
MLRS	-10%	Competitive prototype
		Requirement lowered because of time urgency
		Multiyear procurement, low stretch
Maverick AGM-65A	1%	Total package procurement with low concurrency
		Vigorous testing program
		Low stretch
TOW 2	-4%	Urgent modification program
		Foreign Military Sales
		Low stretch
Sidewinder AIM-9M	10%	Learned from schedule problems in AIM-9L program
		Urgent program, took its lumps in development
		Low stretch
High Growth		
AMRAAM	84%	Prototype showed infeasibility of approach
		High concurrency, rushed testing
		Stretched program, dual sourcing
Phoenix AIM-54C	89%	High concurrency
		Dual-sourced for technical reasons
		Five year qualifying for two years of competition
		Needed funding for next generation
Sparrow AIM-7M	100%	Competitive prototype, low cost growth in development
		Needed funding for next generation
Sidewinder AIM-9L	123%	Crash program
		Dual-sourced for technical reasons
		Production stretch

Fitcher, Arnold, and Allen (1992)

Fitcher et al. (1992) presents a historical perspective of DoD programs schedule performance based on 35 Army, 46 Navy, and 24 Air Force programs from the December 1991 SARs. The purpose of the study is to provide a point estimate and range for the expected schedule duration of future programs by creating probability distributions of past schedule durations within certain intervals. The intervals are: 1) Milestone I – Milestone II, 2) Milestone II - Milestone III, 3) Program start to First flight, 4) Program start to First unit equipped, and 5) Program start IOC. The program interval that most closely relates to our research is the interval from Milestone II to Milestone III.

Although this study in no way tries to predict the schedule duration of a specific interval based on predictor variables, it does provide an ability to check the realism of schedules proposed by the program managers. The probability distributions are compared by service and by intervals to give a range of values as "Most likely" and schedule expectations considered overly optimistic or pessimistic (Fitcher et al., 1992:2). Results from this study show that no marked difference exists among the data from each service and based on the Kolmogorov-Smirnov Goodness-of-Fit test with an alpha of 0.05, all data could be fit to both the normal and the Beta distributions (Fitcher et al., 1992:9).

Also noted, based on the Analysis of Variance (ANOVA) results, is that the only significant difference among the Services (given an alpha level of 0.05) is a longer than average time for Air Force programs compared to the Army and Navy between Milestone II and Milestone III. Service type could prove to be a productive independent variable for our research.

Drezner and Smith (1990)

The results by Drezner and Smith (1990) show the average time to complete a program increases by two years when comparing the timeframes of the 1950s and 60s to the 1970s and 80s. This equates to one year for Phase I and one year for Phase II at a confidence level of 99 and 95 percent, respectively (Drezner and Smith, 1990:9 and 11). However, the authors note that the year the program started fails to capture 90 percent of schedule variance (adj $R^2 = 0.10$) (Drezner and Smith, 1990:9). The results of the study also show the variability of the schedule duration increasing (Drezner and Smith, 1990: vi). Although knowing the duration and variability of schedule is important, understanding what factors make up the duration and affect the variability are imperative.

The researchers of the 1990 RAND study identified 16 potential factors that influence the original schedule and/or subsequent deviations; we list them in Figure 1. Based on statistical analysis of ten programs, the results of the study suggest the following influences on the original schedule estimate: 1) competition and prototyping lengthens schedule and 2) concurrency and adequate funding shortens schedule (Drezner and Smith, 1990:30). Results also suggest the following influences on schedule slips: 1) unstable funding, 2) technical difficulty, 3) external guidance, and 4) external events (Drezner and Smith, 1990:33). Two commonly held hypotheses that prove inconclusive is that longer planning phases incur less slippage, and that cost and schedule growth are interrelated (Drezner and Smith, 1990:40 and 45). However, the authors state, "Our inability to establish these relationships may be due in part to the small database available" (Drezner and Smith, 1990: viii). This study provides a good foundation from

which to proceed forward, but the relatively small data set limits our ability to gain clarity on variables that would be most influential in predicting development schedule.

Factors Affecting Schedule	
Factors Affecting Original Plan	
1. Competition	
2. Concurrency (overlap of effort between development and production phase)	
3. Funding adequacy	
4. Inclusion of prototype phase	
5. If the program's phases were contracted separately	
6. Service priority	(Drezner and Smith, 1990: 21-22)
Factors Affecting Program Deviation	
1. Contractor performance	
2. External events	
3. Funding stability	
4. Major requirements stability	
5. Program manager turnover	(Drezner and Smith, 1990: 23-24)
Factors Affecting Original Plan and/or Program Deviation	
1. External guidance	
2. Single service or joint management	
3. Program complexity	
4. Technical Difficulty	
5. Concept stability (System specification maturity)	(Drezner and Smith, 1990: 23)

Figure 4: Drezner and Smith's Sixteen Schedule Factors

Harmon, Ward, and Palmer (1989)

Harmon et al. (1989) attempt to provide methods for assessing the reasonableness of proposed acquisition schedules. This particular study, consisting of data collected from nine tactical aircraft programs, performs analyses on schedule intervals and provides a schedule assessment tool that spans the period from Full Scale Development (FSD) (now referred to as EMD) start through full-rate production (Harmon et al., 1989:1).

The programs chosen with development occurring from the early 1970s to early 1980s are based on the newness of the program, its importance in historical perspective, and the expected availability of data (Harmon et al., 1989:17). Development program schedules are decomposed into 5 periods: 1) Length of pre-FSD activity, 2) Period from FSD start to first flight, 3) Length of the development flight test program, 4) Early production time, and 5) Total FSD program length (as defined by the period from FSD start to the delivery of the 24th production aircraft) (Harmon et al., 1989:2). Although the researchers refer to these periods as "intervals" they are not mutually exclusive in that certain intervals overlap.

The data is collected from the Office of the Secretary of Defense, military services, contractors, and third parties (studies and databases at IDA, RAND, etc). They obtain cost and technical data from government sources and prime contractors while schedule data is obtained from SARs, contractors, and the services sources (Harmon et al., 1989:17-18).

The program attributes prove to play an important role in explaining variations in interval length. Under the program attribute of the prime contractor, it is estimated that McDonnell Aircraft programs require 15 percent more time than the other four contractor types (Harmon et al, 1989:47). The data also shows that prototype programs require 11 percent less time than programs that do not develop prototypes (Harmon et al, 1989:47). The schedule driver data collected in Figure 5 may be further explored in the building of our model. It should be noted however, not all of these drivers are applicable to our model, since information such as the weight of a completed prototype of low-rate production unit will not be available pre-MS-B.

Candidate Schedule Drivers for Tactical Aircraft

Program Characteristics

1. Military Service
2. Prime Contractor
3. Whether the system was prototyped
4. If the acquisition strategy included contractor teaming
5. If there was separate engine development
6. Number of EMD aircraft built

Aircraft Characteristics

1. Empty weight (lbs.)
2. Combat weight (lbs.)
3. Maximum speed (knots)
4. Thrust to weight ratio at combat weight
5. Mission radius
6. The percentages of titanium and composites used in the airframe structure

(Harmon et al, 1989:19)

Figure 5: Harmon, Ward, and Palmer Schedule Drivers for Tactical Aircraft

Harmon and Ward (1989)

The approach used in this study in many ways parallels that used for the previous study. The data consists of fourteen air-launched missile programs (seven air-to-air and seven air-to-surface systems) that involve substantial developments from the mid-1960s to the late 1980s. Development program schedules are decomposed into 4 periods: 1) Time to first guided launch as measured from FSD start to first guided launch, 2) Length of the development flight test program as measured from the first guided launch to the end of initial operational testing, 3) Early production time as measured from long-lead and full-funding release for the initial production lots to the first production deliveries for those lots, and 4) Program length from first launch as measured by the time from first guided launch to first production delivery.

In the previous study of tactical aircraft, the end of development is stated as the time when 24 aircraft are delivered. Using this methodology for missiles leads to inconsistencies across programs because production rates associated with different types

of missiles vary widely. The researchers decide to use delivery date of the first production missile to mark the end of development (Harmon and Ward, 1989:3).

The data is collected from military services, prime contractors, and third parties (studies and databases at IDA, RAND, etc) with schedule and missiles characteristic data obtained from SARs, numerous government sources, (Harmon and Ward, 1989:8). Collected schedule drivers in Figure 6 may be further explored for the purpose of our research.

Candidate Schedule Drivers for Air-Launched Missiles

Program Characteristics

1. Military Service
2. Prime Contractor
3. Whether the system was prototyped
4. If the system was new or a modification
5. Number of prototype missiles
6. Number of prototype launches
7. Number of development missiles
8. Number of development launches

Missile Characteristics

1. Primary targets
2. Guidance type
3. Length (ft.)
4. Diameter (in.)
5. Total weight (lbs.)
6. Guidance weight
7. Missile Cross-Section (in.^{2})
8. Guidance weight/Cross Section
9. Range (nautical miles)
10. Mach speed
11. Total Impulse (lbs. * sec.)

(Harmon and Ward, 1989:9-10)

Figure 6: Harmon and Ward Schedule Drivers for Air-Launched Missiles

The researchers originally wanted to develop a single equation to predict the interval of FSD defined as the period from FSD start to delivery of the first production missile (Harmon and Ward, 1989:23). Due to the fact that the determinants of time to first launch and time from first launch to first production are vastly different, the

researchers choose the interval between first guided-launch and the first production delivery (Harmon and Ward, 1989:36).

According to the research, time to first launch is a function of technological variables whereas time from first launch to first production is a function of the number of missiles launched in flight test, the rate at which they are launched, the overlap between production start and flight test, and production time (Harmon and Ward, 1989:36).

The researchers believe that the most important determinant of overall development program length is length of the flight test program. Being that flight test duration is determined by the number of test missiles launched and the rate at which test launches are accomplished, it is no surprise that the one program attribute that served most important in determining length of the development effort was the number of missiles launched during flight tests (Harmon and Ward, 1989:13).

Tyson, Nelson, Om, and Palmer (1989)

This study conducted by the IDA examines schedule variances and their causes. The database consists of nine tactical aircraft, nine electronic aircraft, five helicopters, eight other aircraft, 16 air-launched tactical munitions, 18 surface-launched tactical munitions, 10 electronic systems, 10 strategic missiles, and four satellites. The database is divided into four periods: 1960s, early 1970s, late 1970s, and 1980s to compare schedule growth between different timeframes. The results of schedule slippage within the development phase are as follows: 1960s = 46 percent, early 1970s = 24%, late 1970s = 37%, and 1980s = 21% (Tyson et al., 1989:IV-2). The results of schedule slippage within the production phase are as follows: 1960s = 64%, early 1970s = 84%, late 1970s = 69%, and 1980s = 7% (Tyson et al., 1989:IV-2).

The main focus of their study was to determine the effect, if any, on schedule overruns, from: 1) prototyping, 2) competition, 3) multi-year procurement, 4) design-to cost, 5) sole-source procurement and fixed-price development, and 6) contract incentives, variables investigated in previous findings that we have documented. Use of prototyping shows a reduction in the development phase and the overall schedule by 11 and 15 percent, respectively (Tyson et al., 1989:VIII-6 – VIII-7). Competitive programs produce 43 percent more design-schedule growth and 39 percent more production schedule growth, compared to non-competitive programs (Tyson et al., 1989:VII-7). Programs utilizing multiyear-procurement experience seven percent less production schedule growth (Tyson et al., 1989:VI-8). Design-to-cost exhibited development schedule growth of 12 percent and production schedule growth of two percent (Tyson et al., 1989:IX-11). Production schedule growth is reduced by 27 percent when sole-source procurement is used (Tyson et al., 1989:X-7). Under a fixed-price contract strategy, development schedule growth showed a reduction of six percent (Tyson et al., 1989:X-13). It should be noted that no comparison was made between contract incentives and schedules, as that could have been a separate catalyst.

Chapter Summary

In this chapter, we review a multitude of studies that examined various datasets while performing a plethora of statistical procedures all in the pursuit of explaining and predicting schedule duration and variance. It is from these studies that we identify the characteristics that drive acquisition schedules and derive our own list of predictor variables. The accumulation of these predictor variables found throughout the literature review give us a strong foundation from which we can purposefully collect data and

explain our methodology in predicting schedule duration to IOC for defense acquisition programs.

Although the studies reviewed in this chapter differ in the number of programs, the source of data, and methodologies used, they prove beneficial in providing insight into the methodology and predictor variables needed for our research. From past studies, we identified many reasons of schedule growth, schedule variance, and schedule estimating relationships that we wish to investigate as they may be applicable to creating our database and building our regression model. However, it must be noted that not all of the identified variables and relationships may be available in the form of SAR data, thus we now begin the process to manage what information we do have available to us in the SAR. Furthermore, we now develop a foundation from which to begin the methodology for predicting most schedule duration to IOC for defense acquisition programs. The following chapter seeks to addresses the methodology in detail.

III. Methodology

Chapter Overview

This chapter explains the procedures we use to conduct our research. First, we discuss the data source to include its limitations and the process to select and compile the data. Second, define our response variable as it relates to our research question and objectives. Next, we discuss our search for predictor variables, and define candidate predictor variables. We then discuss using preliminary data analysis for the model. Lastly, we discuss the application of a multiple regression analysis, which serves as the statistical cornerstone for predicting a realistic schedule duration for a given acquisitions program.

Database

As mentioned in the previous chapter, Monaco and White (2005, 2006) used a database that had been built and modified over the years by students at AFIT. Because this database is at least 11 years old at this point, we create and employ an entirely new database. The database we utilize for our research is a database originally built by the RAND Corporation for the Air Force Cost Analysis Agency (AFCAA). The SAR database is populated with SAR data on approximately 330 defense acquisition programs. The said SAR database, which is built electronically using separate Microsoft® Excel sheets per program, is in the format of large portfolios of programs grouped by service. The information housed in this major database includes, but is not limited to, vital cost and schedule data necessary for our study.

The database consists of program SARs dating back to the 1950's. Our research seeks to use all programs that contain SAR data that is relevant, applicable, and available

for our multiple regression model. Programs included in the study will contain Air Force, Army, Navy, and Marine Corps. programs. With respect to program type limitations, a U.S. Government Accountability Office (GAO) study on space programs that was presented before the U.S. Senate on 11 May, 2011 states:

> "Despite decades of significant investment, most of the Department of Defense's (DoD) large space acquisition programs have collectively experienced billions of dollars in cost increases, stretched schedules, and increased technical risks. Significant schedule delays of as much as 9 years have resulted in potential capability gaps in missile warning, military communications, and weather monitoring. These problems persist, with other space acquisition programs still facing challenges in meeting their targets and aligning the delivery of assets with appropriate ground and user systems." (GAO, 2011)

Because of the GAO's contemporary findings on extreme cost and schedule growth in space programs despite significant investment to try to mitigate said growth, we choose to exclude space programs from our database to try to preserve the accuracy of our model as it will relate to all other program types.

The SAR database includes program information of all programs, regardless of whether the program was terminated or not. We choose to only include programs that completed IOC. We do this because a cost estimator develops schedule durations based on the idea that the program will be successful and complete IOC. Using successful program data is important because we seek to create regression models that emulate successful programs, which in turn may provide the cost estimator a tool to create a successful development schedule based on past successful program data.

For our study criteria, we consider any program with a "MS-II" labeling to be synonymous with "MS-B" based on each of their respective definitions (Harmon

2012:11). Also, we only include programs that complete the EMD phase up to reported IOC. U.S. Code: Title 10: Section 2432 states:

> "The requirements of this section with respect to a major defense acquisition program shall cease to apply after 90 percent of the items to be delivered to the United States under the program (shown as the total quantity of items to be purchased under the program in the most recent Selected Acquisition Report) have been delivered or 90 percent of planned expenditures under the program have been made." (US Code, 2004)

When a program meets the above criteria, one last SAR report based on the estimate is submitted. This SAR is the one we use to populate our database. It is necessary to wait until a program completes the EMD phase all the way through to the IOC phase to ensure we capture the actual completion date. This determines the amount of schedule duration we use as our dependent variable in model creation.

Furthermore, because we seek to predict schedule duration from the beginning of MS-B to IOC, a major focus of our research database is to include SARs that contain pre-MS-B data. This is significant in that defense acquisition programs are officially initiated at MS-B, and data collection for programs is highly scrutinized at MS-B and beyond. Unfortunately, program data (funding, schedule, etc.) on a program before it is officially initiated at MS-B is not always as highly scrutinized because it is not officially a program at that point in time, and therefore pre-MS-B data is not always as readily available as post-MS-B data.

All the aforementioned characteristics of a program's SAR serves as strict data entry criteria for creation of our research database. SARs that had all of the characteristics except one was not considered due to the fact that incomplete data on a

program would not be of use in analyzing that data, nor would incomplete data be useful in the multiple regression model we seek to build.

With respect to the SAR database, an iterative process of criteria (Figure 3) was applied to all of the programs in the database as to filter for only programs we could use for our study.

Table 4: Process of Database Filtering

		At Start	Filtered Out from Criteria	Remaining
Criteria 1	*No Space Programs*	330	24	306
Criteria 2	*MS-B to IOC (Months)*	306	80	226
Criteria 3	*% of RDT&E Funding at MS-B Start (BY16)*	226	117	109
Criteria 4	*MS-A to MS-B Duration (Months)*	109	53	56

This filtering process helped us get to the sample size of 56 programs we use for our research. First, we seek to filter out all space programs, as previously mentioned. Second, we seek to use only programs that give us both MS-B and IOC dates, as this will be our response variable. Next we look for RDT&E funding data as it relates to a percentage of total RDT&E funds allocation at MS-B. This idea comes from Brown et al. (2015) and Unger (2001) who found that front-loading a program's RDT&E funding has a correlation to lessened schedule growth. In order for a percentage to be calculated, there needs to be at least one year prior to MS-B of RDT&E funding data. This criteria filtered out the most programs simply due to the fact that many of the final SARs in the database only showed funding data at MS-B and thereafter. Finally, the last criteria also relates back to the Brown et al. (2015) and Unger (2001) findings in that calculating RDT&E funds percentage allocated at MS-B can best be captured from clearly defined MS-A start, finish, and funding data.

Selected Acquisition Report (SAR) Data

The SAR contains an array of major defense acquisition program data from all military services. At a minimum, this data includes schedule, cost, budget, and performance characteristics of defense programs. Although the criteria for a program to be classified as ACAT I change over time, the programs within the SAR consistently represent programs of high interest to the government. The SAR data can include both classified and unclassified information. For security reasons, we only include unclassified information in our database.

As seen in the literature review, SAR data is commonly used to conduct research on both schedule and cost growth. Even though the government has made improvements in both quality and consistency of information within SAR data, there are still many weaknesses with respect to data collection and reporting that get manifested in missing or incomplete data (Hough, 1992:v). Inconsistencies exist due to the fact that guidelines change over time, and specific details vary from program to program leading to complications with interprogram comparisons (Hough, 1992:4). Even with the traditional limitations associated with SAR data, it still remains a logical source of data for our research due to the wide range of information it has on programs that are of high interest to the government.

Response Variable

This research utilizes a multiple regression approach to predicting program schedule duration. We express the multiple regression response as time duration in months for our modeling database, although the predicted response more than likely will have remaining time expressed as a decimal of a month. Therefore for usage of the

model, we suggest rounding the predicted response to a whole number. The overall time duration in months starts at EMD, through the Production and Deployment, LRIP, and IOT&E phases, and concludes at IOC.

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

- *MS-B to IOC (Months) [Regression Output]*
    - This variable states the actual time it took from MS-B to IOC for a given program. This data is unavailable to the cost estimator at the time they are developing a cost estimate.

The accuracy of the "MS-B to IOC (Months)" response variable will be dependent on the strength of the predictor variables associated to it in the multiple regression model.

Search for Predictors of Schedule Duration

Our past studies discussed in Chapter II identify possible predictor variables as they relate to our research. To be of value in the application of cost estimating, it is imperative the explanatory (independent) variables are both understandable and available to the cost estimator when the program office begins the schedule estimate as part of the cost estimate.

A variable that is predictive yet confusing, or unavailable to the estimator, is essentially worthless if it cannot be communicated to an audience, or understood by another user. For this reason, we create models consisting of clearly defined variables that the cost estimator is confident in. This produces a model that has utility and is easily defendable. In the search for predictors, we do not mandate a causal relationship to the response variable, but the independent variable must exhibit some logical link to the response variable that the cost estimator can easily understand. Furthermore, along with

the cost estimator being able to understand the independent variable, the data associated with the independent variable should accessible in common reporting standards.

Predictor Variables

The candidate explanatory variables used in the multiple regression model to predict schedule duration come exclusively from the SAR database. With strict data entry criteria applied to creating our research database, the predictor variables found have some logical link to the program, and should be readily available to the cost estimator on a given SAR deliverable. Our final regression model, described next in Chapter IV, only includes those predictor variables that prove statistically significant at $\alpha=0.05$ level of significance. Next, we list and describe the predictor variables considered for inclusion in the multiple regression model.

All of these were found across all 56 programs of our modified SAR database: [Note: only relevant categorical variables are listed here if that particular type of program was in the database. Since there are no ships in our research database, there is no explanatory variable listed as Ship. The same can be said for Tank, etc.]

- *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 (BY16) – Continuous Variable*
    - This variable is based on simply 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 current base year at the time of this research (BY16).

- *% of RDT&E Funding at MS-B Start (BY16) – 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 % was derived from were all standardized into the current base year at the time of this research (BY16).

- *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 includes 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 et al, 2011:246).

- *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.

- *MS-B Start Year – Continuous Variable*

- This variable addresses the year in which MS-B started. Much like the "1985 or Later for MS-B Start" predictor variable shown above, the actual year in which MS-B started has the probability of significance on the schedule duration of a program.

- *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 criteria 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 that 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 Products).

- *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.

- *Contractor (Name of Defense Contractor(s)) – Binary Variable*
    - This variable identifies the name of the lead defense contractor for a given weapons system program. If the effort on a program involved more than one contractor, a variable was created with all named contractors sharing that variable.
- *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.

Validation Pool

Once all data is gathered across all 56 programs, we randomly select 20 percent of the 56 programs to serve as our validation pool. This means we build our multiple regression model with the data of 45 programs, while the other 11 completed programs' data is used to test the multiple regression model against for accuracy of output.

Exploratory Data Analysis

Inherent in building a valid, statistically significant multiple regression model is the application of various statistical techniques that can further help us to create the most robust model possible. It should be noted that a test for independence is not part of our exploratory data analysis. Due to the fact we use only one SAR to obtain data for any one program, we assume independence is met, although we have no way of statistically testing this assumption.

Variance Inflation Factors

One of the first analyses done in the exploration of the data involves looking at the variance inflation factors (VIF) scores. We seek to display and analyze the VIF scores of any predictor variables that prove to be statistically significant. Variance inflation is a consequence of multicollinearity and the VIF scores are a common way for detecting

such a relationship (Stine, 1995). When an independent variable is nearly a linear combination of other independent variables in the model, the affected estimates are unstable and exhibit high standard errors. This is due to a linear dependency between two or more independent variables where the value of one predictor is dependent upon another (Stine, 1995). “A VIF of 10 suggests that it is large enough to indicate a problem” (Stine, 1995).

Cook’s Distance Test

To make sure there are no overly influential data points that are creating skewed outputs in our model, we look to Cook’s Distance test, commonly referred to as “Cook’s D” (Cook, 1977). Cook's D is a commonly used estimate of the influence of data point(s) when performing a regression analysis. Cook's D can be used in several ways: to indicate data points that are particularly worth investigating for validity, to indicate regions of a space where it would be good to be able to obtain more data points, or even removing data points that appear to be overly influential in our regression model. All of these uses of Cook’s D should be applied on a case by case basis. For the purpose of our research, we look to Cook’s D to check for any program data that is overly influential to our model using JMP®. Typically, we are justified in removing a data point when the Cook’s D value is 0.5 or greater. A Cook’s D value that is greater than 0.5 indicates that an influential data point exists (Neter et al., 1996:381).

Studentized Residuals

In juxtaposition with Cook’s D, which looks for influential data points, we also look at the histogram of studentized residuals to identify potential outliers. If we detect any potential outliers in the studentized residual histogram, we further explore on an

individual basis to see if the outlier should be kept in the regression model. This anomaly could indicate a data transcription issue, a rare occurrence, an atypical program, or for a host of reasons that cannot be explained. For purposes of our research, we consider any program whose studentized residual is either 3 standard deviations above or below then standard normal distribution's mean of zero. This is in keeping assumption a normal distribution of the residuals, which we discuss next.

Shapiro-Wilk's Test

Any multiple regression model that we ultimately settle upon must have its model residuals pass the assumption of being normally distributed and possessing constant variance. These two assumptions are needed to satisfy/maintain the validity of the models' p-values.

The Shapiro-Wilk's (S-W) goodness of fit test (Neter et al., 1996: 111) addresses the normality assumption. The S-W test is a way to statistically determine whether a random sample comes from a normal distribution or not. We use a threshold of $\alpha = 0.05$ to conduct this test. The null hypothesis for the S-W is that the model residuals possess a normal distribution. The alternative hypothesis is that they do not. If the p-value for the S-W is larger than 0.05, then we can satisfy the assumption of normality for the data in our model.

Breusch-Pagan Test

Following this, we test our final model assumption of constant variance of the error term using the Breusch-Pagan (B-P) test (Neter et al., 1996:239). This test for constant variance in a regression model is used with the purpose of identifying whether heteroscedasticity is present in the model or not. Heteroscedasticity refers to the

circumstance in which the variance of an explanatory variable is not constant (unequal) across the range of values of a different variable that predicts it. In order to have the most robust regression model possible, having as close to equal constant variance as possible is most advantageous.

We conduct the B-P test for our research using Microsoft® Excel after obtaining data inputs via JMP®. In order to pass the assumption of constant variance using the B-P test, the p-value output from the test must be above 0.05. Similar to the S-W test, the null hypothesis states that our assumption with respect to the model's residuals (for the B-P test, this is constant variance) holds.

Stepwise Regression

We use the process of stepwise regression to assist us in determining which explanatory variables prove both individually predictive as well collectively predictive. The stepwise function in JMP® gives give us a preliminary regression model to work with, and all of the aforementioned exploratory data analysis methods will be conducted following the output of a preliminary multiple regression model. Thus, chronologically, a multiple regression analysis is run first, but it will not be used or considered significant until all exploratory data analysis methods are conducted and satisfied.

We use the mixed direction within the stepwise regression in lieu of the forward and backwards option. The purpose of this is to optimize both the fitting routine as well as to prevent carrying non-predictive variables once more predictive variables are added to the preliminary regression model.

We use a p-value threshold of 0.05 for an explanatory variable to enter the model as well as a value of 0.05 to leave the model. That is, for an initial explanatory variable

to be entered into the regression model, it must have a p-value of less 0.05. Once within the model, if it ever reaches a p-value of greater than 0.05 (once other explanatory variables are included) stepwise then removes this variable.

Multiple Regression Analysis

The last step of our model building process involves the ultimate finalization of our multiple regression model once stepwise has produced an initial model and we have ascertained there are no issues (and tested that) with respect to multicollinearity (VIF scores), influential data points (Cook's D), outliers (studentized residuals) and satisfied the assumption of normality (S-W test) and constant variance (B-P). The structure of the finalized model reflects the standard linear multiple regression equation (1).

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \ldots + \beta_k X_{ki} + \varepsilon_i \text{ (McClave et al., 2001:557)} \qquad (1)$$

Where:

Y_i - Outcome of Dependent Variable (response) for i^{th} experimental/sample unit
X_i - Level of Independent (predictor) variable for i^{th} experimental/sample unit
$\beta_0 + \beta_1 X_i$ - Linear /systematic relation between Y_i and X_i (conditional mean)
β_0 - Mean of Y when X=0 (Y-intercept)
β_1 - Change in mean of Y when X increases by 1 (slope)
ε_i - Random error term

In this finalization, we also seek to make sure that the final model is statistically significant at our chosen experimentwise error rate of 0.05, but that we also ensure each explanatory variable is significant with respect to its respective comparisonwise error rate. This later requirement is necessary such that we don't erroneously violate the experimentwise error rate for the overall model's F test while conducting multiple t-tests for the individual model parameters.

This later step requires us to adopt a procedure to control for the overall Type I error rate by adopting a familywise error rate procedure. For this research, we utilize the Bonferroni Correction (Bonferroni, 2015). The application of this corrective measure is an adjustment made to P-values when several dependent or independent statistical tests are being performed simultaneously on a single data set. To perform a Bonferroni Correction, divide the P-value (α) by the number of comparisons being made (m). The output of this will give us α_c which will be the threshold by which each P-value must be less than to remain in the model. If an independent variable gets removed from the regression model by way of the Bonferroni Correction, a new iteration will be conducted with a new value for the number of comparisons (m), which will in turn create a newly calculated α_c threshold. This is an iterative process and can take multiple iterations, but it serves as strong conservative measure to avoid the potential of having a lot of spurious positives in the testing of the data set (Bonferroni, 2015). The only way an independent variable can remain in the model is if it fails the threshold by only a small margin (and small is contextual), and in each case an analysis will be conducted on the importance of keeping said independent variable in the model.

The multiple regression model can only be considered complete and valid upon passing all phases of the exploratory data analysis. If at any point the multiple regression model fails any phase of the exploratory data analysis, proper remedial measures will be taken, and an iterative process will take place until significant results are present in a model.

Descriptive performance measures we utilize for the multiple regression analysis are the R^2 and Adjusted R^2 outputs. The R^2 is a statistical measure of how close the data

fit to the regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression. An R^2 of 0 indicates that the model explains none of the variability of the response data around its mean, while R^2 values closer to 1 have a much stronger explanation. Adjusted R^2 has been adjusted for the number of predictors in the model. The Adjusted R^2 increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance. Adjusted R^2 ensures we do not arbitrarily add variables to the model that are not predictive. Adjusted R^2 ensures we do not arbitrarily add variables to the model that are not predictive.

$$\text{Adjusted } R^2 = 1 - \left[\frac{(n-1)}{n-(k+1)}\right](1-R^2) \quad \text{(McClave et al., 2001:557)} \qquad (2)$$

Where:

n= the number of data points
k= the number of independent variables in the model

As seen in (2), the value of Adjusted R^2 decreases when we add additional variables to the model. If the added variable increases the explained variance noted in R^2, Adjusted R^2 increases. Therefore, this counterbalance ensures that we add variables whose predictability warrants the additional complexity of the model (McClave et al., 2001:557).

Validation of Multiple Regression Model

The final model is tested against our validation pool. Our model, consisting of 45 programs, is tested against our validation pool of 11 programs. Two measures of this validation takes place. First, we compute the Absolute Percent Error (APE) for each program and then determine the Mean and Median Absolute Percent Error (MAPE and

MdAPE) for both sets of data. The APE equals | actual MS-B to IOC (months) minus predicted MS-B to IOC (months) | divided by actual MS-B to IOC (months). The MdAPE and MAPE measure the average prediction accuracy of each regression model's outputs. We then compare the MdAPE and MAPE to check that they behave the same as both the larger sample and smaller sample should exude similar percent errors in their characteristics.

Once the MdAPE and MAPE are compared, we then construct a predicted by actuals bivariate plot to compare the regression line of both graphs. Once again, we check to see that both the main model and model built from the validation pool behave in a similar fashion. If our main model holds up against both measures of the MdAPE and MAPE comparison, as well as the bivariate plot, we can combine the original 45 programs with the 11 programs of the validation pool, and we can create a finalized model using all 56 programs.

Chapter Summary

We use the results of our literature review as a foundation for our analytical process. This chapter details our foundation by describing our research methodology. We explore the use of SAR data, describe our process of data collection, and explain our creation of predictor variables that provide a link to the response variable. We provide reasoning for the use of our methodology and provide a detailed explanation of the exploratory data analysis conducted on the data to further help us create the most robust model possible. We drive forward into the next chapter to introduce the results of our model building process.

IV. Results and Analysis

Chapter Overview

This chapter provides the results from the methodology outlined in Chapter III. First, using our model pool of 45 defense acquisition programs, we run a preliminary multiple regression analysis in JMP® using the stepwise function. Second, we conduct our data analysis techniques on the preliminary multiple regression model as a means to validate assumptions about the model, which gives us a final model. Next, using our finalized multiple regression model, we measure, compare, and discuss our statistically significant predictor variables. Then we discuss the explanatory power of our model overall using the R^2 and Adjusted R^2 values. Finally, we judge the performance of our finalized multiple regression model by testing it against our validation pool of 11 programs, along with measuring the validation performance as it relates to raw output accuracy with respect to the MdAPE and MAPE range.

Preliminary Multiple Regression Model

Applying the stepwise function in JMP® to our data on 45 programs, we arrive at the output displayed in Figure 7. This figure shows us that our preliminary model appears to display many characteristics that would help us to predict schedule duration to IOC for a given program.

In our preliminary model, we note the presence of many of the predictor variables that were detailed in Chapter III. Also detailed in Chapter III, the Bonferroni Correction can be applied to the model, as a conservative measure to avoid any potential spurious positives from testing the model.

Response MS-B to IOC (Months) [Regression Output]

Summary of Fit

RSquare	0.465725
RSquare Adj	0.412298
Root Mean Square Error	24.84426
Mean of Response	75.13333
Observations (or Sum Wgts)	45

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	21521.715	5380.43	8.7170
Error	40	24689.485	617.24	Prob > F
C. Total	44	46211.200		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	82.071139	8.655695	9.48	<.0001*
RDT&E $ (M) at MS-B Start (BY16)	0.0077162	0.002693	2.87	0.0066*
% of RDT&E Funding at MS-B Start (BY16)	-86.70413	24.15955	-3.59	0.0009*
Modification	-19.34527	8.989336	-2.15	0.0375*
1985 or Later for MS-B Start	19.58554	7.635146	2.57	0.0142*

Figure 7: Preliminary Multiple Regression Model Output

Applying the Bonferroni Correction to this preliminary model requires us to calculate α_c by accounting for all of the independent variables, except the intercept, and dividing that number by the P-value. For this study being exploratory in nature, we use an α of 0.05 for this step. This gives us 0.05/4 = 0.0125 as our α_c. By this standard, the *Modification* predictor variable would be removed from the model because it is above the α_c threshold by 0.0125.

As mentioned in Chapter III, the only way an independent variable can remain in the model is if it fails the α_c threshold by only a small margin, and in each case an analysis will be conducted on the importance of keeping said independent variable in the model. In this case, *Modification* is above the α_c by a relatively small enough margin that we see value in investigating its potential to remain in the model. At first, we notice the

Modification predictor variable has a negative value as the coefficient. Because our *Modification* predictor variable is binary, this tells us that when a defense program is characterized as modification that we can expect it to truncate the schedule duration output.

Outside of what the data suggests, this seems reasonable to us because a modification to a weapons system that has already been developed and operational could indeed have a higher probability of a quicker duration to IOC, as compared with a new program that is being developed and tested for the first time. Based on our investigation, we decide to keep the *Modification* in the model for the reasonableness of its predictive nature. Also, because the Bonferroni Correction is defined as an additional conservative measure, we acknowledge the conservatism associated with it, but choose not to apply it in this case (Bonferroni, 2015).

At this point, because the Bonferroni Correction was not applied, we consider this our preliminary model. Working with this preliminary model, we now apply the previously described data analysis techniques as a means to seek validation of our assumptions in the model.

Validating Model Assumptions

The multiple regression model assumptions will be considered validated upon passing all data analysis techniques that are applied to it. If at any time the multiple regression model fails any of the data analysis techniques, proper measures will be documented and executed, and an iterative process will take place until the deficiency is remediated.

Please note that if the model fails a particular data analysis technique, we will stop in the phase in which it failed, and it will be dealt with and re-analyzed in the current phase it is in. This purpose serves to show detailed continuity in the process, without restarting the entire process for each failure potentially encountered.

Variance Inflation Factors

With respect to our preliminary model in Figure 8, we see that the VIF scores are all well under the value of 2. While all VIF scores are under 2, all of the VIF scores remaining are actually in the lower range, closer to that of a VIF score of 1. The analysis of this tells us that there is no consequence of multicollinearity present in the preliminary model. By this, there is no linear dependency between two or more independent variables where the value of one predictor is dependent upon another (Stine, 1995). With all of the preliminary model's VIF scores passing the test, we move onto the Cook's D test.

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	82.071139	8.655695	9.48	<.0001*	.
RDT&E $ (M) at MS-B Start (BY16)	0.0077162	0.002693	2.87	0.0066*	1.0568113
% of RDT&E Funding at MS-B Start (BY16)	-86.70413	24.15955	-3.59	0.0009*	1.0062185
Modification	-19.34527	8.989336	-2.15	0.0375*	1.0880836
1985 or Later for MS-B Start	19.58554	7.635146	2.57	0.0142*	1.036805

Figure 8: Preliminary Model VIF Scores

Cook's Distance Test

Looking to our preliminary model in Figure 7, we now conduct the Cook's D test on the data of our 45 programs to test the influence of data point(s) when performing our multiple regression analysis (Cook, 1977). As noted before, a Cook's D value that is greater than 0.5 indicates that an influential data point exists (Neter et al., 1996:381). Influential data points may be removed from the data set upon investigation, justification,

and documentation outlining the process by which the decision was made to remove the data point. Figure 9 displays the Cook's D for our model.

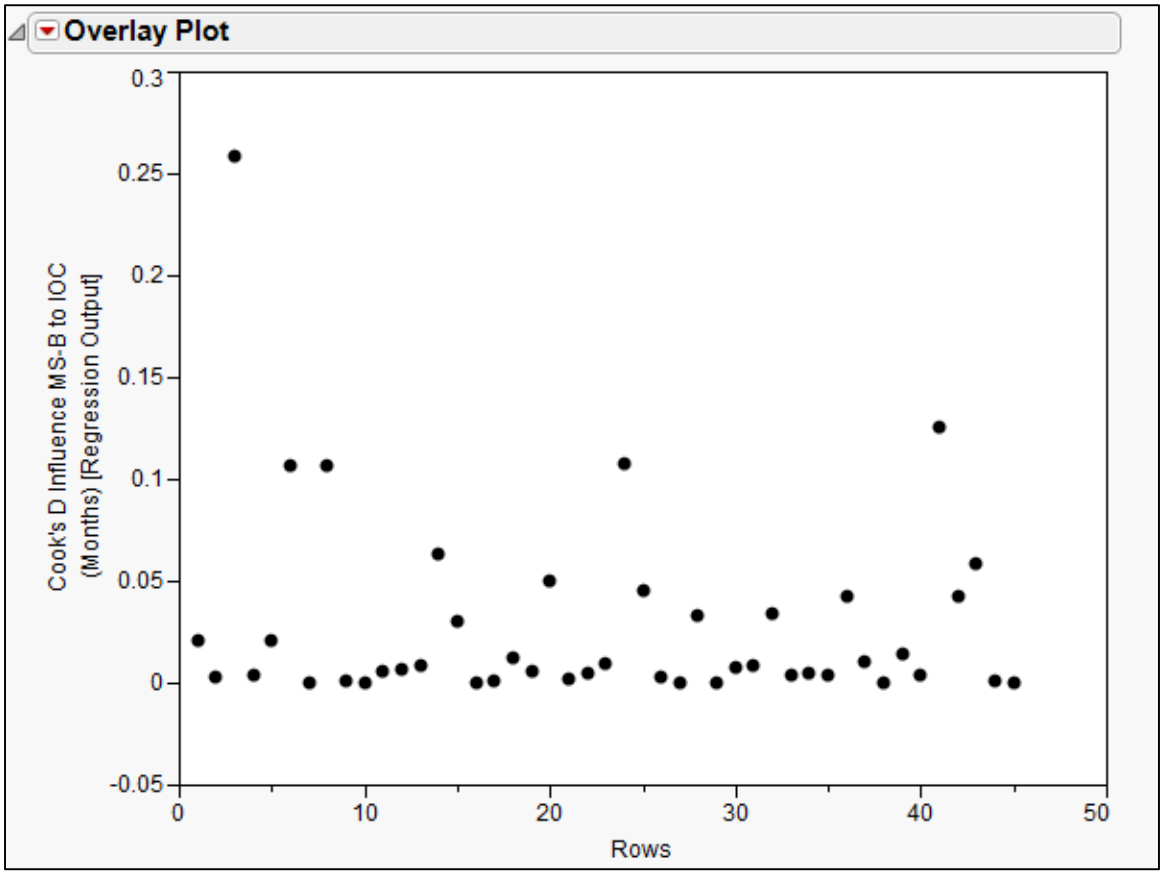


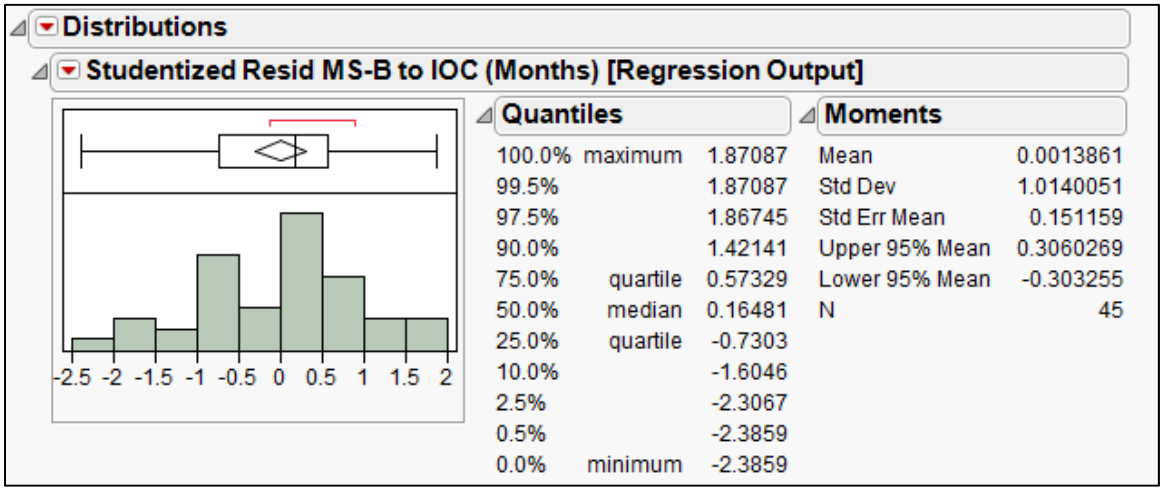
Figure 9: Display of Cook's D Plot

Our Cook's D plot displays all data points below the previously defined 0.5 threshold. This means the Cook's D test conducted on the 45 programs shows no data points that are influential on our preliminary multiple regression model. With the Cook's D test showing no influential data points (Figure 9), we now look for potential outliers in the data set.

Studentized Residuals

We generate a histogram (Figure 10) of the studentized residuals to look for potential outliers in the data. Since all studentized residuals lie between 3 and –3 on this

graph, there appears no outliers for us to worry about. Given the relatively normal distribution shape, we expect that when we test the assumption of normality on the non-studentized residuals via the S-W test, that this assumption will be validated.



Quantiles		
100.0%	maximum	1.87087
99.5%		1.87087
97.5%		1.86745
90.0%		1.42141
75.0%	quartile	0.57329
50.0%	median	0.16481
25.0%	quartile	-0.7303
10.0%		-1.6046
2.5%		-2.3067
0.5%		-2.3859
0.0%	minimum	-2.3859

Moments	
Mean	0.0013861
Std Dev	1.0140051
Std Err Mean	0.151159
Upper 95% Mean	0.3060269
Lower 95% Mean	-0.303255
N	45

Figure 10: Studentized Residuals

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

As detailed earlier in Chapter III, the S-W goodness of fit test (Neter et al., 1996: 111) assesses the assumption of normality with respect to the residuals of the multiple regression model, while the B-P assesses the assumption of constant variance.

Since both Figure 11 and Table 4 indicate P-values greater than our established criteria of 0.05, we fail to reject the null hypothesis for either test. [Note: Figure 12 displays the sum of squares for regression (SSR) that we need for the B-P test conducted in Excel.] Therefore, we conclude our multiple regression model passes both model residual assumptions.

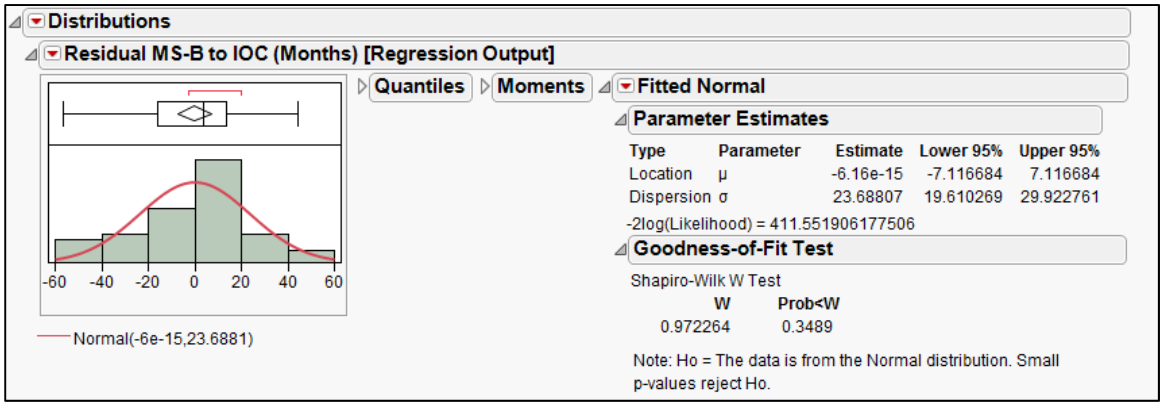
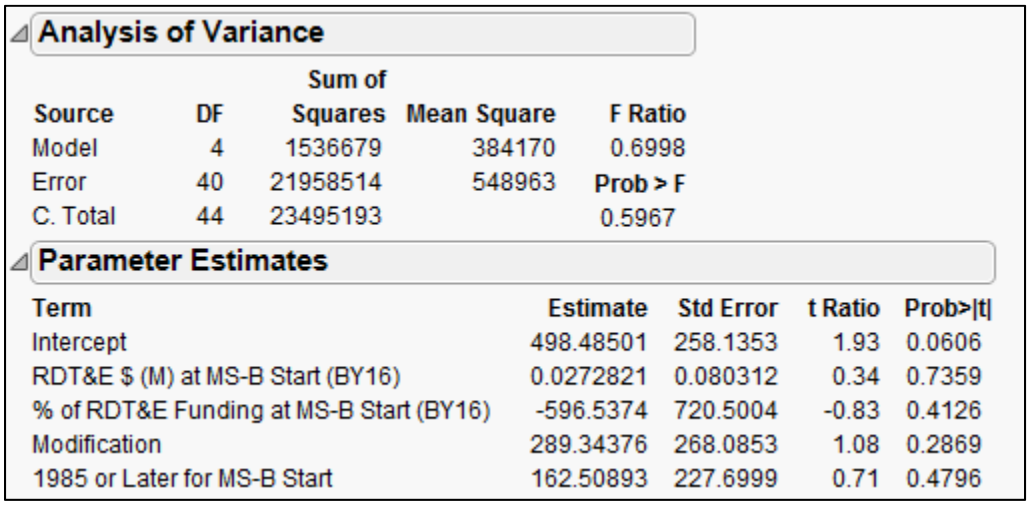


Figure 11: Shapiro-Wilk's Test



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	1536679	384170	0.6998
Error	40	21958514	548963	Prob > F
C. Total	44	23495193		0.5967

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	498.48501	258.1353	1.93	0.0606
RDT&E $ (M) at MS-B Start (BY16)	0.0272821	0.080312	0.34	0.7359
% of RDT&E Funding at MS-B Start (BY16)	-596.5374	720.5004	-0.83	0.4126
Modification	289.34376	268.0853	1.08	0.2869
1985 or Later for MS-B Start	162.50893	227.6999	0.71	0.4796

Figure 12: ANOVA Output

Table 5: Breusch-Pagan Test Results

			B-P Test Statistic	P-Value
Sample Size	45			
Model Degrees of Freedom	4		2.552431679	0.635272337
SSE	24689.485			
SSR	1536679			

Validation of Assumptions

The statistical tests that were performed on our preliminary regression model from Figure 7 were done so to try to validate our previously mentioned assumptions about the model. Because all statistical tests were passed to validate our assumptions, we

can consider our preliminary regression model from Figure 7 as our non-preliminary, main model for the continuing purpose of our research.

Analysis of Predictor Variables

Our statistically significant predictor variables tell us individually something about themselves outside of their presence in the validated model. The parameter estimates for our model are displayed in Figure 13. We deep dive each individual predictor variable, and discuss the estimates associated with each.

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta	VIF
Intercept	82.071139	8.655695	9.48	<.0001*	0	.
RDT&E $ (M) at MS-B Start (BY16)	0.0077162	0.002693	2.87	0.0066*	0.340422	1.0568113
% of RDT&E Funding at MS-B Start (BY16)	-86.70413	24.15955	-3.59	0.0009*	-0.41605	1.0062185
Modification	-19.34527	8.989336	-2.15	0.0375*	-0.25944	1.0880836
1985 or Later for MS-B Start	19.58554	7.635146	2.57	0.0142*	0.301869	1.036805

Figure 13: Parameter Estimates for Predictor Variables

The predictor variables that are statistically significant in the validated model are listed along with an analysis of each next:

- *RDT&E $(M) at MS-B Start (BY16) – Continuous Variable*
    - The parameter estimate associated with this variable is 0.00772 which would be multiplied by the raw amount of RDT&E funding in BY16 dollars (millions) allocated to the program at MS-B start. It should be noted that even if the overall RDT&E funding outlay of a program is uniformly distributed for the entirety of its RDT&E expenditures, the variable only looks at purely the raw amount of BY16 dollars at MS-B start. Perhaps the 0.00772 estimate output for this variable is associated with the idea that the raw amount of RDT&E dollars (BY16) that are present in a program at the time of MS-B start is related to "technology" or even "technology maturity". While there is no way to prove that within the scope of our research, the multiple of 0.0072 seems to account for anticipated complexity of a system in predicting schedule duration from MS-B to IOC, as more raw money for this variable is an additive factor to schedule output.

- *% of RDT&E Funding at MS-B Start (BY16) – Continuous Variable*

- This variable is the strongest variable with respect to its Standard Beta (weight) in Figure 13. The parameter estimate of -86.704 for the predictor variable suggests the greater the % of RDT&E funding that has been allocated at MS-B, the greater the decrease of schedule duration from MS-B to IOC. The idea of this variable can be linked back to the purpose of the technology maturation and risk reduction (TMRR) phase, which occurs immediately prior to MS-B. According to DoDI 5000.02 (USD(AT&L), 2015), the purpose of TMRR "is to reduce technology, engineering, integration, and life-cycle risk" before program initiation. Based on this definition, we theorize that increasing the % of RDT&E funding prior to program initiation (MS-B start) is synonymous with increasing technology maturity and reducing risk prior to program entry at MS-B. Our finding is supported by the Unger et al (2004) study, which finds that program RDT&E budgets that can be fit with an increasingly right-skewed Weibull distribution encounter less schedule growth, on average.

- *Modification – Binary Variable*
  - This variable is -19.345 which means that when the program being analyzed by our regression model is characterized as a modification that it should take away from the overall schedule duration output of the model. Because our literature included many studies that alluded to the idea that a higher probability of cost and schedule problems raise when programs start with technologies at low readiness levels, a modification having a shortening effect on schedule output seems reasonable to us because a modification to a weapons system that has already been developed and operational could indeed have a higher probability of a quicker duration to IOC, as compared with a new program.

- *1985 or Later for MS-B Start – Binary Variable*
  - Every program schedule created contemporarily will use a "1" for the binary applicability of this predictor variable. The parameter estimate of 19.586 suggests that programs after 1985 will actually add time to the schedule duration of a program. For some, this may seem counterintuitive in that it could be argues that technology gets better as time goes on, and therefore program schedule should be shorter as time goes on because of this. On the other hand, systems are becoming much more complex as time goes on, and the technical maturity of a weapons system that needs to meet the demands of 21st century warfare could actually take longer with time due to the high-level of complexity. Originally, this variable was discovered by Brown et al. (2015) in reference to the President's Blue Ribbon Commission on Defense (commonly called the Packard Commission) and the subsequent acquisition reforms. In the current environment of tight

> budgets, heightened acquisition reform, and weapons systems being more complex than ever before, it seems completely reasonable that a program with an MS-B start after 1985 would add time to the MS-B to IOC schedule duration.

The predictor variables that were found to be statistically significant in our regression model all give strong contributions to the method of predicting schedule duration to IOC. At the most basic level, our predictor variables have a strong logical link to predicting schedule duration for a program. Furthermore, all of the predictor variables in our model are accessible and available to the cost estimator via data that can be found in the SAR of a program.

Performance of Multiple Regression Model

We judge the performance of our multiple regression model using the R^2 and Adjusted R^2 values as shown in Figure 17. An R^2 of 0.465 indicates that the model explains almost 50 percent of the variability associated with predicting time duration from MS-B to IOC. Brown et al. (2015) acknowledges that cost and schedule estimates are rarely clairvoyant, particularly in the early stages of a program. Because we seek to predict program schedule duration to IOC at the very beginning of a program's life cycle (program initiation), an R^2 of 0.465 can actually be considered strong when taking into account the on-going documented problems with schedule growth from our literature review.

Our model's Adjusted R^2 is 0.412 and we can also gauge the Adjusted R^2 to be relatively strong, based on the true lack of clarity regarding a program's schedule, especially in the very beginning of a complex weapons system acquisition. Since Adjusted R^2 ensures we do not arbitrarily add variables to the model that are not

predictive, the distinction should be made that we can highlight the R^2 of this model to a cost estimator or decision maker as a descriptor of the model's robustness, but the Adjusted R^2 of this model is the value that should be focused on when making decisions.

Response MS-B to IOC (Months) [Regression Output]

Summary of Fit

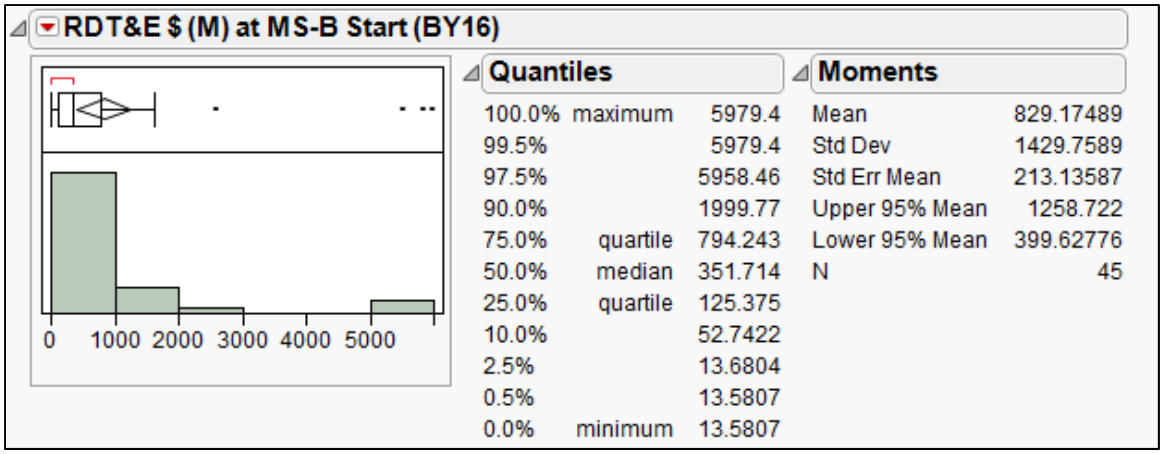
RSquare	0.465725
RSquare Adj	0.412298
Root Mean Square Error	24.84426
Mean of Response	75.13333
Observations (or Sum Wgts)	45

Figure 14: Model R^2 and Adjusted R^2 Values

Validation of Multiple Regression Model

As a matter of testing predictive ability of our finalized model, we compare the accuracy of our fitted multiple regression model against programs with information from the research validation database. But prior to this, we first mention the range of our explanatory variables for which this model can be used. This is to prevent model extrapolation.

We present histograms of the range of values we can input for the continuous variables, X_1 and X_2 in Figures 15 and 16. In our histograms of the *RDT&E $(M) at MS-B Start (BY16)* variable and % *of RDT&E Funding at MS-B Start (BY16)* variable, we see our ranges are between $13.581M and $5,979.4M (BY16) , and 1.09 percent and 59.2 percent, respectively.



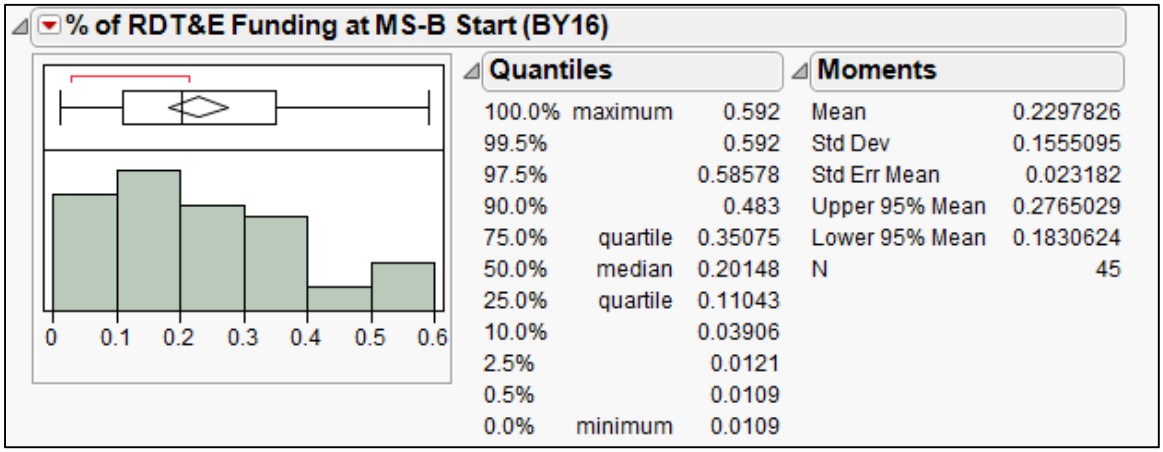
Quantiles

100.0%	maximum	5979.4
99.5%		5979.4
97.5%		5958.46
90.0%		1999.77
75.0%	quartile	794.243
50.0%	median	351.714
25.0%	quartile	125.375
10.0%		52.7422
2.5%		13.6804
0.5%		13.5807
0.0%	minimum	13.5807

Moments

Mean	829.17489
Std Dev	1429.7589
Std Err Mean	213.13587
Upper 95% Mean	1258.722
Lower 95% Mean	399.62776
N	45

Figure 15: RDT&E $ (M) at MS-B Start (BY16) Quantiles



Quantiles

100.0%	maximum	0.592
99.5%		0.592
97.5%		0.58578
90.0%		0.483
75.0%	quartile	0.35075
50.0%	median	0.20148
25.0%	quartile	0.11043
10.0%		0.03906
2.5%		0.0121
0.5%		0.0109
0.0%	minimum	0.0109

Moments

Mean	0.2297826
Std Dev	0.1555095
Std Err Mean	0.023182
Upper 95% Mean	0.2765029
Lower 95% Mean	0.1830624
N	45

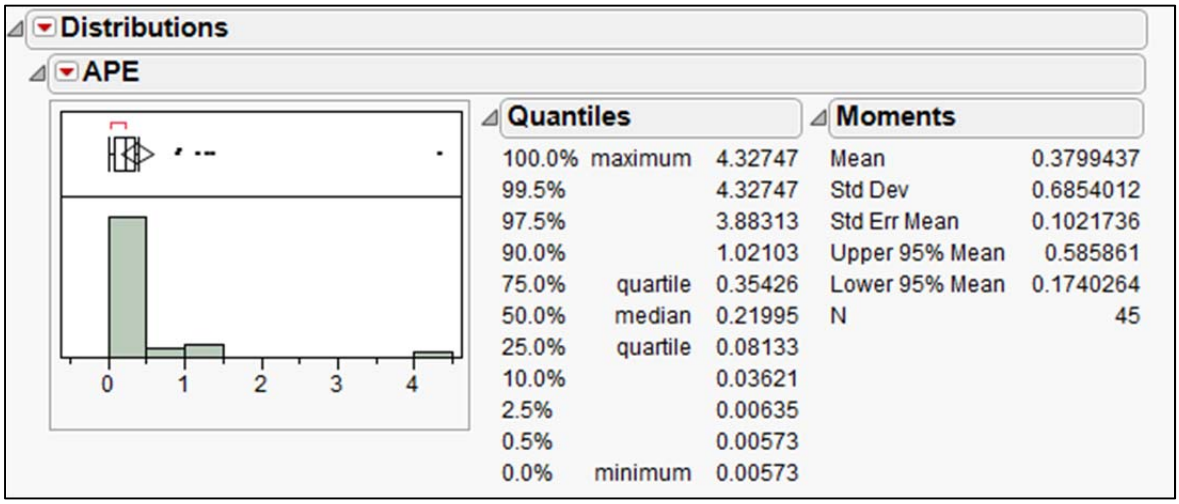
Figure 16: % RDT&E Funding at MS-B Start (BY16) Quantiles

Next, we proceed to look the MAPE and MdAPE associated with our model built from the 45 original programs. We also look at the MAPE and MdAPE of the 11 programs from the validation pool, and compare both models.

In Figure 17, we note that the MAPE is 0.379 and the MdAPE is 0.219 for our model built from 45 programs. Of the 45 programs, we also notice 6 outliers in the histogram. Of the outlier subset, three are electronics programs, two are missiles programs, and one is a bomber program. What we can note about the electronics and missiles programs is that they had a relatively low time frame to reach IOC. In the missiles programs, one was a modification, and two of the electronics programs were

modifications. The lone bomber program (A-10) experienced a relatively low time to IOC, probably because the first generation of this aircraft was relatively low in complexity.

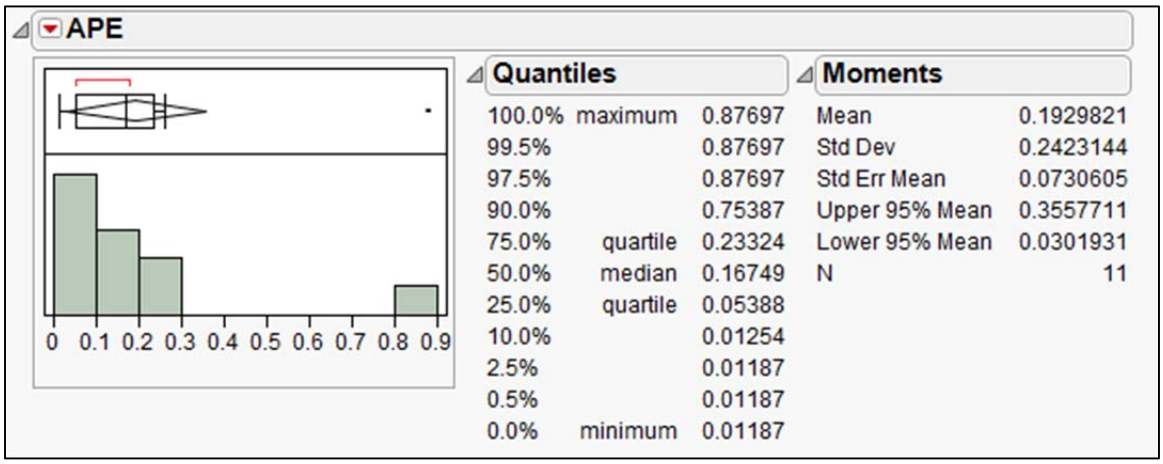
In Figure 18, we note that the MAPE is 0.193 and the MdAPE is 0.167 for our validation pool of 11 programs. Of the 11 programs, we also notice one outlier in the histogram. The outlier is a modification program to a bomber aircraft (B1-B).



Quantiles		
100.0%	maximum	4.32747
99.5%		4.32747
97.5%		3.88313
90.0%		1.02103
75.0%	quartile	0.35426
50.0%	median	0.21995
25.0%	quartile	0.08133
10.0%		0.03621
2.5%		0.00635
0.5%		0.00573
0.0%	minimum	0.00573

Moments	
Mean	0.3799437
Std Dev	0.6854012
Std Err Mean	0.1021736
Upper 95% Mean	0.585861
Lower 95% Mean	0.1740264
N	45

Figure 17: MdAPE and MAPE of Final Model



Quantiles		
100.0%	maximum	0.87697
99.5%		0.87697
97.5%		0.87697
90.0%		0.75387
75.0%	quartile	0.23324
50.0%	median	0.16749
25.0%	quartile	0.05388
10.0%		0.01254
2.5%		0.01187
0.5%		0.01187
0.0%	minimum	0.01187

Moments	
Mean	0.1929821
Std Dev	0.2423144
Std Err Mean	0.0730605
Upper 95% Mean	0.3557711
Lower 95% Mean	0.0301931
N	11

Figure 18: MdAPE and MAPE of Validation Pool

Due to the great disparity of sample sizes in each of the respective APE and MAPE outputs, along with the lack of normality from the distribution of the outputs, we

look to the MdAPEs as the much more representative numbers for comparing our sample outputs. With the final model having a MdAPE of 0.219 and the validation pool model having a MdAPE of 0.167 we can see that they are not far off from one another. This gives us some confidence in saying the two models are comparable. However, we can gain more confidence if we look to a comparison of actual by predicted plots of both the final model and validation pool; this will gives a visual representation of the predictive power of each of the models with their respective sample sizes.

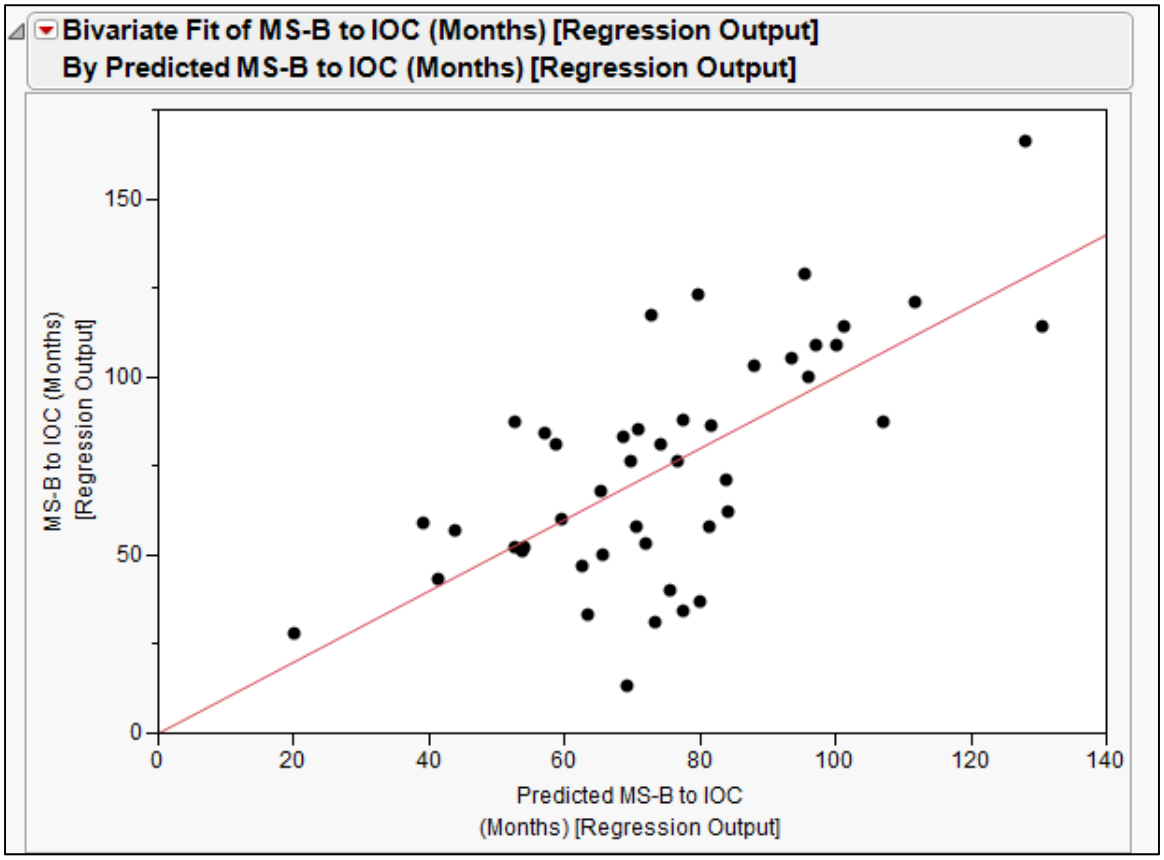


Figure 19: Bivariate Plot of Model with 45 Programs

In Figure 19 we see a relatively good fitted line to our 45 data points. Please note that while our line does intersect some of the data points, there still tend to be many points that are away from the line, but none so egregious that it causes concern.

Therefore, for the intended purpose of predicting schedule duration to IOC, this fitted line seems to satisfy our intended use of the model. While confidence intervals are not applied to this fitted line on the graph, we speculate that a decision maker would be inclined to adjust the predicted schedule duration output based on their experience and knowledge of a program. Next, we look to the bivariate plot of our validation pool in Figure 20.

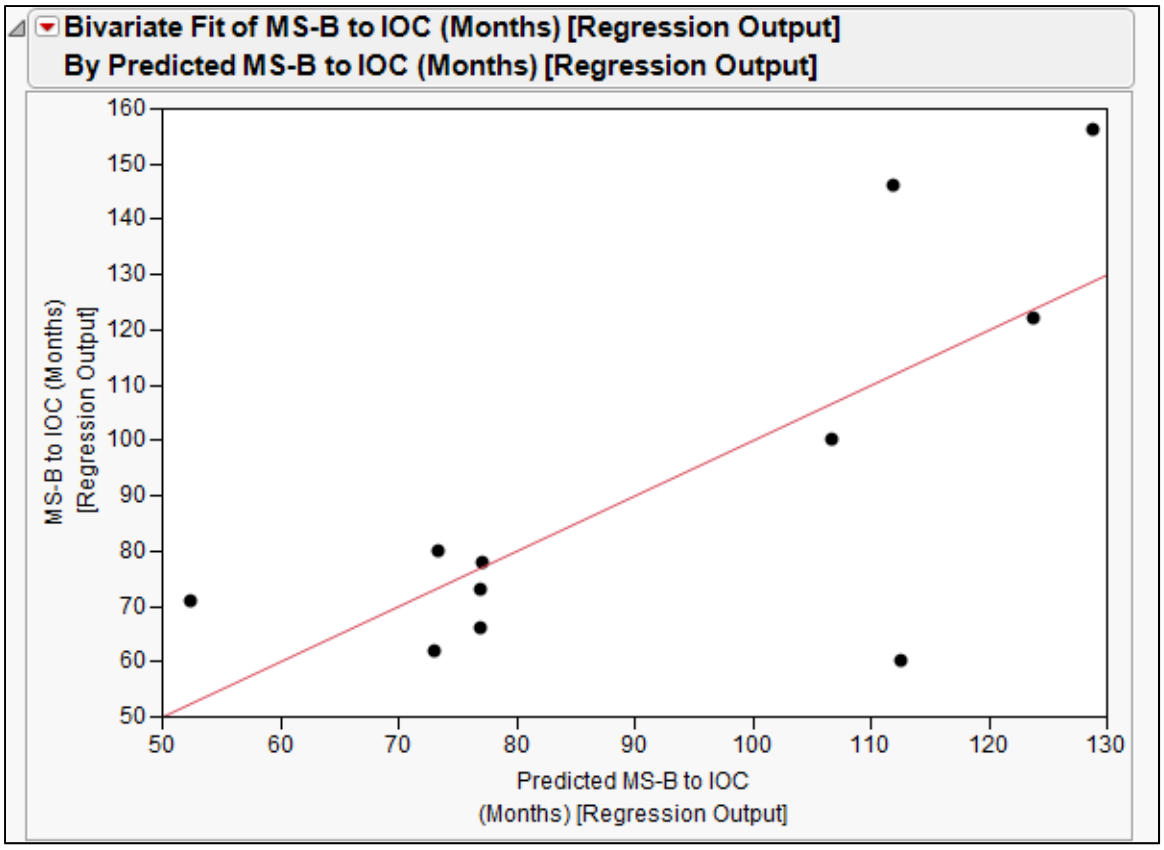


Figure 20: Bivariate Plot of Validation Pool with 11 Programs

In Figure 20, we see our fitted line to the validation pool of 11 programs. Notice the line is relatively close to seven of the 11 data points, while other points seem to be a little bit further away. While confidence intervals are not applied to this fitted line on the graph, we propose that as many as two additional data points could make it into the predicted output range of the displayed 11 data points.

Overall, due to comparison of the MdAPE and MAPE of both the final model and validation pool, along with comparison of the predicted by fitted bivariate outputs, we can consider out model valid. Therefore, finally, we compile all the data from the final model and the validation pool to just update variable parameters, and this becomes our complete final model, thus concluding the validation part of your analysis. When our 45 programs are combined with the 11 from the validation pool, our final model using 56 programs is displayed in Figure 21.

Summary of Fit

RSquare	0.428807
RSquare Adj	0.384008
Root Mean Square Error	26.03015
Mean of Response	78.48214
Observations (or Sum Wgts)	56

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	25941.974	6485.49	9.5717
Error	51	34556.008	677.57	Prob > F
C. Total	55	60497.982		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta	VIF
Intercept	81.938131	7.406529	11.06	<.0001*	0	.
RDT&E $ (M) at MS-B Start (BY16)	0.0079896	0.002745	2.91	0.0053*	0.321926	1.0924985
% of RDT&E Funding at MS-B Start (BY16)	-89.33287	22.93536	-3.89	0.0003*	-0.41735	1.0251399
Modification	-18.34103	8.498096	-2.16	0.0356*	-0.24163	1.1191264
1985 or Later for MS-B Start	24.792533	7.171663	3.46	0.0011*	0.374979	1.0505097

Figure 21: Final Model with all 56 Programs

We can see that final model in Figure 21 with all 56 programs, when compared to the preliminary model of 45 programs in Figure 7, holds much of the same validity when compared to one another. The R^2 and Adjusted R^2 are still somewhat relatively the same with only a minor change in both, the intercept only went down by one month, all of the independent variables remain significant when the Bonferroni Correction is not applied as a conservative measure, and the VIF and Stand Betas also hold their same

characteristics. This is our full and complete final multiple regression model with the data of all 56 defense acquisitions programs.

Chapter Summary

In this chapter, we create a preliminary multiple regression model, validate model assumptions, validate the model, and report the results of our finalized multiple regression model for predicting schedule duration of a program from MS-B to IOC. We explain some of our findings to include statistical testing applied to the regression model built. We continue with a separate, in-depth analysis for each of the predictor variables that were found to be statistically significant in the final model. We further solidify our belief that our multiple regression model is robust, parsimonious, and statistically sound through judgement of our performance measures. Lastly, in our validation of the model, we bring all 56 programs together to create a finalized model multiple regression model that is statistically significant. In the next chapter, we conclude our research and present some broad discussions and meaning to our analysis.

V. Conclusions and Recommendations

Chapter Overview

This chapter summarizes the quantitatively-focused method in our research that is driven by the data of past weapons systems. The major finding in our research was a statistically significant multiple regression model, which may be used to predict schedule duration to IOC for a program. First, we revisit our initial research questions to validate that our research accomplished its intended goal. Additionally, we review the limitations of findings, identify areas for future research, and conclude by summarizing the significance of this research.

Research Questions Answered

1 – ***Can we accurately predict what the schedule duration of a defense acquisition program should be, from MS-B to IOC, using a mathematical model?***

With respect to the final model we created and the available data we were able to gather, the answer is yes. Schedule duration to IOC output can be given for any program that has available data inputs necessary to populate the model. All of the data necessary for the continuous and binary variables can be gathered from the SAR in any given program at MS-B, such that our model is statistically significant in predicting MS-B to IOC schedule duration using only the data available up to MS-B start.

2 – ***Can we statistically show that some independent variables are stronger than others when it comes to predicting a future program's schedule duration?***

As outlined in Chapter IV when we deep dive into the analysis of each predictor variable and its effect on the multiple regression model, the answer to this research question is yes. In the analysis of the predictor variables, each predictor variable's

parameter estimate gives us a foundation from which we can statistically infer that some variables are stronger than others as far as predictive properties are concerned for our model.

Our two strongest predictors of schedule duration were the *1985 or Later at MS-B Start* variable and the *% RDT&E Funding at MS-B Start (BY16)* variable. Of the two, the *% RDT&E Funding at MS-B Start (BY16)* variable is the stronger predictor, as noted by Standard Beta outputs from JMP®. In Figure 22, we display a pie chart showing the percentage contribution for each Standard Beta as it is associated with its independent variable.

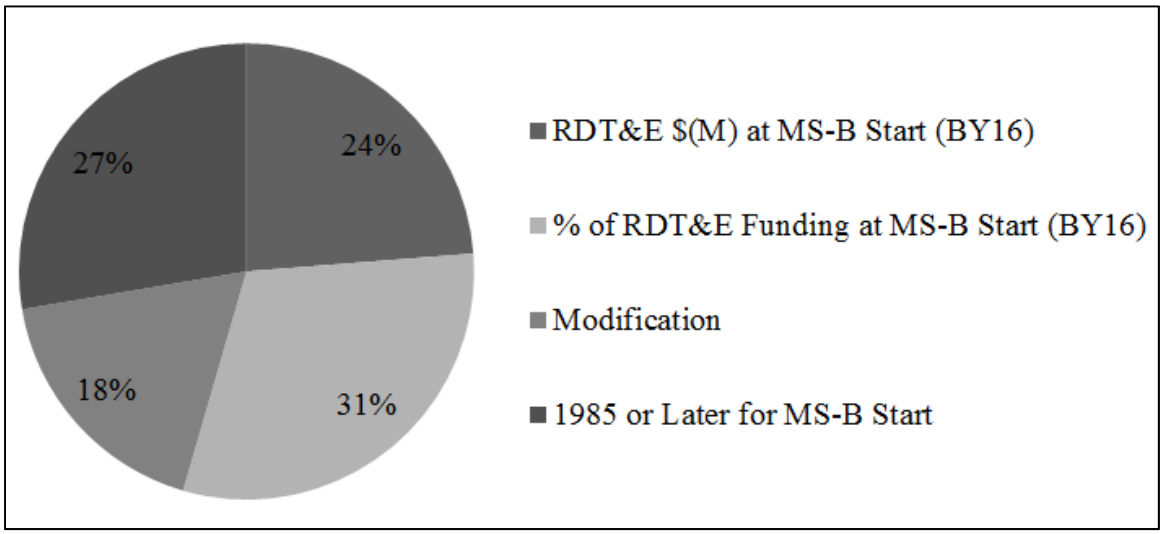


Figure 22: Pie Chart of Standard Betas

Another predictor variable that showed strength in predictive ability was a post-1985 MS-B start date. This perhaps accounted for the increasing complexity of weapons systems over time, along with effects of the Packard Commission, and serves as an additive factor to the model. Modification programs were seen as to have a postulated schedule efficiency associated with them, considering the binary variable took away from the schedule ration output. Finally, we note that is statistically significance in the sheer

amount of money a program has at MS-B start. It produces a slight additive factor, as to account for cost growth seen in programs that heavier amounts of funding by way of their complexity.

Findings

The biggest finding was the *% RDT&E Funding at MS-B Start (BY16)* variable. This is significant because it directly relates to previous studies from Brown et al. (2015) and Unger (2001), which found a correlation between front loading RDT&E funding and minimized schedule growth. Our *% RDT&E Funding at MS-B Start (BY16)* variable sought to identify the percentage of a front-loaded RDT&E funding profile at MS-B for a program if it existed. For the purpose of our model, those programs that apply a more front-loaded RDT&E funding profile at MS-B, they should see a lessened schedule duration to IOC, as the statistics suggests.

Furthermore, there were no significant findings in the planned concurrency of a schedule. Planned concurrency did not show to be statistically significant in a positive or negative impact to the model. Along with this, the planned quantity of a program's units, specific contractor, and a program that planned prototypes all were not shown to be statistically significant in predicting schedule duration to IOC. Finally, the model was shown to be service-agnostic, meaning there was no distinguishable schedule characteristics in which service the program was for, according to our model.

Limitations

We recognize several major limitations of this research, and that could potentially limit the application of it in the greater cost analysis community. First and foremost, we

must recognize that our model can only be as good as the data that was used for it. The availability of pre-MS-B data was perhaps the strongest limitation we encountered. The scarcity of available pre-MS-B data was a major proxy that led us to source 56 programs for our research database from the original SAR database. Of the data that we originally set out to gather based on our highlighted findings in the literature review, we had to further narrow the scope of data collection because much of that data was simply not available us in the SAR database we used.

Perhaps the most important pre-MS-B information that was not available to us in the SAR database was the TRL of a program. Many of the studies in our literature review tested the idea that schedule growth in a program has a strong correlation to the technological maturity necessary for the program going into MS-B. More pre-MS-B data available on programs would be necessary to highlight more predictor variables, and of that necessary data that was not available, prior studies particularly suggest that the TRL of a program could potentially have been a statistically significant predictor variable in our model.

Our finalized model was developed using data from 56 different programs, which. The total amount of programs used could be another limitation in our model. First, we must acknowledge that some of the studies in our literature review used less than 56 programs in their studies, but other studies in our literature review used more than 56 programs. Comparison of studies on programs may not be completely analogous due to the scope of program types used in a previous study, or the availability of their respective data. However, we can postulate a sensitivity analysis on the number of programs used in

our model, with the idea that the model could maybe have been more robust if more programs were used that had the available and necessary data we used in our model.

Finally, we look to the potential accuracy limitations associated with the final model output. While we did answer our first research question of creating a statistically significant model that can accurately predict schedule duration to IOC, we do recognize that rarely does one model fit perfectly for all of its future intended uses. Schedule duration to IOC output can be given for any program type that has available data inputs necessary to populate the model, but we must also address the adjustment factor for the *Modification* predictor variable. In using the *Modification* predictor variable for a bomber program, it can be hypothesized that our model's schedule duration output for a modification program may be slightly more precise in its accuracy when compared to using the model for non-modification programs.

All of the stated limitations in the research can, in some way, be tied back to the availability of the necessary data available to us in our model building process. Our model shows that various types of pre-MS-B activity can be predictive characteristics in predicting schedule duration of a program. The idea that pre-MS-B data could help predict other aspects of a program, such as cost or production, should not completely be ignored. Perhaps this could suggest a future push to require pre-MS-B data collection of future programs, should that program experience any pre-MS-B activity.

Recommendations for Future Research

Recommendations for future research encourage the exploration and use of the original SAR database, as well as our modified research database of 56 programs. Whereas our research is the first to explore predicting schedule duration using this

methodology, we acknowledge that follow-on research and other methodologies used to predict schedule duration can be of great value to the great cost community; especially when we take into account the upcoming should-schedule initiative being put into place by Secretary James. We highly encourage further exploration into program schedule research, as it can directly or indirectly support the new should-schedule initiative. For instance:

- Collect more SAR data to further populate our research database with more pre-MS-B data from programs, and then perform the same methodology we used to build a multiple regression model that predicts schedule duration from MS-B to IOC. Perhaps more predictor variables could be identified in the model, along with new R^2 and Adjusted R^2 values.

- Employing the SAR database to create numerous multiple regression models that do not explicitly rely on pre-MS-B data. The numerous models would be used to tell us predicted times for various other points in a program's schedule, i.e. time from MS-C to IOC, time from PDR to CDR, etc.

- Perform sensitivity analyses on our model by varying each of the independent variable inputs.

- Add a competition variable to our database and determine if this variable adds to the predictability of our model.

- On a live defense acquisition program, use linear and non-linear programming to optimize the timeliness of a program's schedule with respect to the program's already predicted schedule. The linear and non-linear programming model(s) could serve as the actual should-schedule value(s) for the program.

Chapter Summary

Accurately predicting the most realistic schedule for a program, especially at the official initiation of a program, is an extremely difficult task considering the inherent risk and uncertainties that are present in the early stages of a program. Programs that decide to use an unnecessarily lengthy schedule as a program strategy run the risk of delaying the level of technological advancement that may be critical to national safety. However,

accelerated program schedules increase the risks of unscheduled delays and expensive rework and retooling costs, especially if a problem is found later in the accelerated program schedule (Drezner and Smith, 1990: iii). Our research creates a mitigation tool against both elongated and aggressive schedule durations by quantitatively predicting a schedule duration outcome based on historical program data.

The most noted difference between our research and previous research on schedule is our use of a multiple regression analysis to predict the schedule duration of a defense acquisition program. We recommend the use of a multiple regression model as a top-level management tool to aid in identifying the duration of a program schedule at program initiation. We believe the previously untapped resource of using a multiple regression analysis to predict schedule duration provides a valuable tool, and merits a great deal of utility, to both cost estimators and decision makers alike.

In addition to providing predicted schedule duration as an output, our model could add value by serving as a cross-check to a program that already has created a schedule estimate to IOC. Furthermore, our model also provides the cost estimator with a schedule benchmark that they can use to try to employ operational efficiencies in a program as to try to deliver a program's capability quicker than what the historical data suggests; application in this form directly supports Secretary James' should-schedule strategy.

Appendix A: List of Acronyms

ACAT – Acquisition Category
AFCAA – Air Force Cost Analysis Agency
AFIT – Air Force Institute of Technology
AFRL – Air Force Research Laboratory
ANOVA – Analysis of Variance
APE – Absolute Percent Error
B-P – Breusch-Pagan Test
BY – Base Year
CDR – Critical Design Review
CER – Cost Estimating Relationship
CPR – Cost Performance Report
DAU – Defense Acquisition University
DoD – Department of Defense
EMD – Engineering and Manufacturing Development
FSD – Full Scale Development
FUE – First Unit Equipped
GAO – Government Accountability Office
IDA – Institute for Defense Analyses
IOC – Initial Operating Capability
IOT&E – Initial Operational Test and Evaluation
LRIP – Low Rate Initial Production
MAPE – Mean Absolute Percent Error
MdAPE – Median Absolute Percent Error
MS – Milestone
NASA – National Aeronautics and Space Administration
OSD – Office of the Secretary of Defense
P&D – Production and Deployment
PCA – Production Contract Award
PDR – Preliminary Design Review
RAND – Research and Development Corporation
RDT&E – Research Development Test & Evaluation
SAR – Selected Acquisition Report
SECAF – Secretary of the Air Force
S-W – Shapiro-Wilk's Test
SECM – Systems Engineering Concept Tool and Method
SME – Subject Matter Expert
TMRR – Technology Maturation and Risk Reduction
TRL – Technology Readiness Level
VIF – Variance Inflation Factor

Appendix B: 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 C: List of 56 Programs in Research Database

1	A-10 (SAR date at MS-B, March 1973) (BY70)
2	AWACS (SAR date at MS-B, July 1970) (BY70)
3	C-17 (SAR date at MS-B, Dec 1985) (BY81)
4	F-22 (SAR date at MS-B, Aug 1991) (BY85)
5	AH-64 (SAR date at MS-B, Dec 1976) (BY72)
6	B-1B CMUP-Computer (SAR date at MS-B, May 1996) (BY95)
7	C-5 RERP (SAR date at MS-B, Dec 2001) (BY00)
8	F-15 (SAR date at MS-B, Jan 1970) (BY70)
9	B-1B JDAM (SAR date at MS-B, Mar 1995) (BY95)
10	KC-135R (SAR date at MS-B, Jan 1980) (BY81)
11	B-1B Defense System Upgrade (SAR date at MS-B, Jun 1997) (BY96)
12	FA-18 A/B (SAR date at MS-B, Jan 1976) (BY75)
13	AV-8B Harrier (SAR date at MS-B, Aug 1980) (BY79)
14	S-3A (SAR date at MS-B, Aug 1969) (BY68)
15	P-8 Poseidon (SAR date at MS-B, June 2004) (BY04)
16	V-22 Osprey (SAR date at MS-B, May 1986) (BY84)
17	E-2C Hawkeye (SAR date at MS-B, May 1969) (BY68)
18	F-35 JSF (SAR date at MS-B, Oct 2001) (BY94)
19	CH-47D Chinook (SAR date at MS-B, June 1976) (BY75)
20	E-8A JSTARS (SAR date at MS-B, Sept 1985) (BY83)
21	AGM-65A Missile (SAR date at MS-B, July 1968) (BY68)
22	ALCM Missile (SAR date at MS-B, Oct 1977) (BY77)
23	AMRAAM Missile (SAR date at MS-B, Dec 1981) (BY78)
24	CSRL (SAR date at MS-B, June 1983) (BY82)
25	JASSM Missile (SAR at MS-B, Nov 1998) (BY95)
26	JDAM (SAR at MS-B, Oct 1995) (BY93)
27	JPATS T-6A (SAR at MS-B, Feb 1996) (BY95)
28	MARK XV Identification FoF (SAR at MS-B, Feb 1989) (BY82)
29	Microwave Landing System [MLS] (SAR at MS-B, Aug 1988) (BY82)
30	OTH-B (SAR at MS-B, June 1982) (BY82)
31	LGM-118 Peacekeeper (SAR at MS-B, Sept 1979) (BY82)
32	GBU-39 SDB-I (SAR at MS-B, Oct 2003) (BY01)
33	MGM-134 SICBM (SAR at MS-B, Dec 1986) (BY84)
34	SRAM-II Missile (SAR at MS-B, Aug 1987) (BY83)
35	National Aerospace System (SAR at MS-B, July 1995) (BY90)
36	ADS (SAR at MS-B, Sep 2004) (BY05)
37	AGM-88 HARM (SAR at MS-B, Aug 1978) (BY78)
38	AIM-9X Block 1 (SAT at MS-B, Dec 1996) (BY92)

39	**AN/BSY-1 (SAR at MS-B, Dec 1983) (BY84)**
40	**ASDS (SAR at MS-B, Sep 1994) (BY03)**
41	**COBRA Judy Replacement (SAR at MS-B, Dec 2003) (BY03)**
42	**Harpoon Missile (SAR at MS-B, June 1971) (BY70)**
43	**JSOW-BL (SAR at MS-B, June 1992) (BY90)**
44	**NATBMD (SAR at MS-B, Sep 1997) (BY94)**
45	**NMT (SAR at MS-B, May 2007) (BY02)**
46	**SH-60B (SAR at MS-B, Jan 1978) (BY76)**
47	**UGM-96A Trident I Missile (SAR at MS-B, Aug 1974) (BY74)**
48	**SSN 774 (Virginia Class Sub) (SAR at MS-B, Jan 1996) (BY94)**
49	**T-45TS (SAR at MS-B, Oct 1984)(BY1984)**
50	**YAL-1 (SAR at MS-B, March, 1996) (BY97)**
51	**UGM-109 Tomahawk (SAR at MS-B, Jan 1977) (BY77)**
52	**SSBN 726 Sub (SAR at MS-B, July 1974) (BY74)**
53	**AGM-114A Hellfire Missile (SAR at MS-B, Oct 1976) (BY75)**
54	**OH-58D Helicopter (SAR at MS-B, Sep 1981) (BY82)**
55	**AAWS-M Javelin (SAR at MS-B, June 1989) (BY90)**
56	**SSN 21 Sub (SAR at MS-B, Jan 1989) (BY85)**

Appendix D: Contractors in Research Database

C1	**Fairchild**
C2	**Rockwell**
C3	**McDonnel Douglas**
C4	**General Dynamics**
C5	**Lockheed Martin**
C6	**Lockheed and Boeing**
C7	**Beech Aircraft Corp**
C8	**Boeing**
C9	**Boeing and Bell**
C10	**Northrop Grumman**
C11	**Hughes**
C12	**Hughes and Raytheon**
C13	**Allied Corp**
C14	**Textron**
C15	**General Electric**
C16	**Texas Instruments**
C17	**IBM and GE**
C18	**Raytheon**
C19	**IBM Federal Systems**
C20	**Bell-Textron**
C21	**Newport**

Appendix E: Data for 56 Programs in Research Database

#	MS-A to MS-B Duration (Months)	MS-B to IOC (Months) [Regression Output]	Quantity Expected at MS-B	RDT&E $ (M) at MS-B Start (BY16)	% of *RDT&E* Funding at MS-B Start (BY16)	Modification	Prototype	Concurrency Planned	1985 or Later for MS-B Start	Year	Air Force	Navy	Army	Marine Corps	Aircraft	Fighter Program	Bomber Program	Helo Program	Cargo Plane Program	Tanker Program	Electronic Warfare Pro	Trainer Plane Progran	Missile Program	Electronic System Proj	Submarine Program	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	ACAT I
1	27	33	729	439.98	0.2522	0	1	1	1970	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2	20	80	42	375.39	0.0827	0	1	0	1970	0	1	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
3	52	109	210	433.14	0.0564	0	1	1	1985	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
4	58	166	648	5497.8	0.1831	0	1	1	1991	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
5	35	85	472	933.18	0.2114	0	1	1	1976	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
6	37	60	103	8.8387	0.0245	1	0	1	1996	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	22	146	108	76.611	0.0377	1	1	1	2001	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	13	58	729	401.55	0.0442	0	0	1	1970	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
9	23	37	221	14.245	0.0281	1	0	1	1995	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	25	52	334	26.186	0.1183	1	0	1	1980	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	50	123	60	13.581	0.0314	1	0	1	1997	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	49	86	800	75.472	0.0109	0	1	1	1976	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
13	53	60	336	1040.6	0.3511	0	1	1	1980	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
14	12	50	193	1262.9	0.3004	0	0	1	1969	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
15	29	114	108	163.84	0.0189	0	0	1	2004	1	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
16	37	156	913	874.74	0.1803	0	1	1	1986	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
17	36	28	28	655.31	0.5505	1	1	1	1969	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
18	59	114	2852	5839.8	0.185	0	1	1	2001	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
19	14	71	361	17.767	0.0462	1	1	1	1976	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	40	129	10	280.56	0.097	0	1	1	1985	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
21	25	43	17000	351.71	0.5	0	0	1	1968	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
22	40	62	3424	595.08	0.2422	0	1	1	1977	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	34	81	24335	448.59	0.1284	0	1	0	1981	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
24	12	81	104	181.91	0.2848	0	0	0	1983	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	29	58	2400	464.34	0.3977	0	0	0	1998	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	18	31	87496	264.99	0.3504	0	1	1	1995	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	37	76	711	81.733	0.1512	1	1	1	1996	1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
28	55	103	15747	397.71	0.1934	0	1	0	1989	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
29	55	47	10111	62	0.2323	1	0	0	1988	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
30	81	66	7	14.805	0.0184	0	1	0	1982	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1
31	42	87	223	5979.4	0.2445	0	0	0	1979	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	26	34	70	143.82	0.2893	0	1	0	2003	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	36	121	0	2582.3	0.1133	0	1	0	1986	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	30	100	1633	176.41	0.0802	0	1	0	1987	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	36	59	53	96.89	0.5067	1	1	0	1995	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	16	52	15	490.21	0.592	0	0	0	2004	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	70	57	13754	338.72	0.4717	0	1	1	1978	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
38	24	83	10000	126.5	0.1671	1	0	1	1996	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
39	21	40	0	563.76	0.1228	0	0	1	1983	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
40	22	105	5	53.104	0.096	0	1	0	1994	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
41	16	109	1	118	0.0627	0	0	1	2003	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
42	15	73	4200	62.91	0.0497	0	0	1	1971	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	36	71	8800	128.39	0.2159	0	1	0	1992	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
44	59	84	1500	1161.1	0.3936	1	0	0	1997	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45	43	53	305	313.39	0.3698	0	1	1	1997	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
46	45	68	204	614.24	0.2451	0	1	1	1978	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
47	32	62	340	1611.4	0.12	0	0	0	1974	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48	27	88	1	1356.2	0.3979	0	0	0	1996	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
49	25	78	12	73.141	0.0666	0	0	0	1984	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
50	28	122	0	1428.1	0.3158	0	1	0	2003	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
51	35	113	1062	572.13	0.2015	1	1	1	1997	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
52	27	87	10	1107.2	0.4375	0	0	0	1974	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
53	28	117	24600	124.25	0.1165	0	1	1	1976	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
54	22	51	578	52.2	0.1076	1	1	0	1981	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
55	34	76	70550	272.1	0.3114	0	1	0	1989	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
56	44	100	12	2846.4	0.3986	0	0	0	1989	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

Appendix F: Response and Predictor Variables

Response Variable:

- *MS-B to IOC (Months) [Regression Output]*

Predictor Variables:

- *MS-A to MS-B Duration (Months) – Continuous Variable*
- *Quantity Expected at MS-B – Continuous Variable*
- *RDT&E $ (M) at MS-B Start (BY16) – Continuous Variable*
- *% of RDT&E Funding at MS-B Start (BY16) – Continuous Variable*
- *Modification – Binary Variable*
- *Prototype – Binary Variable*
- *Concurrency Planned – Binary Variable*
- *1985 or Later for MS-B Start – Binary Variable*
- *MS-B Start Year – Continuous Variable*
- *Air Force – Binary Variable*
- *Navy – Binary Variable*
- *Army – Binary Variable*
- *Marine Corps – Binary Variable*
- *Aircraft – Binary Variable*
- *Fighter Program – Binary Variable*
- *Bomber Program – Binary Variable*
- *Helo Program – Binary Variable*
- *Cargo Plane Program – Binary Variable*

- *Tanker Program – Binary Variable*
- *Electronic Warfare Program – Binary Variable*
- *Trainer Plane Program – Binary Variable*
- *Missile Program – Binary Variable*
- *Electronic System Program – Binary Variable*
- *Submarine Program – Binary Variable*
- *Contractor (Name of Defense Contractor(s)) – Binary Variable*
- *ACAT I – Binary Variable*

Appendix G: Validation Pool of 11 Programs

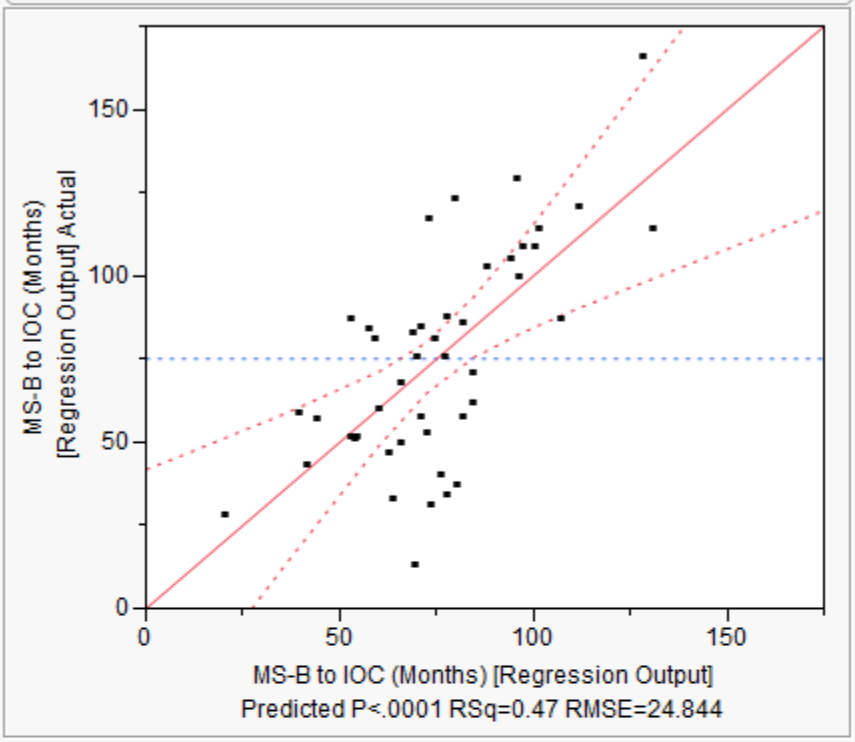
1	AWACS (SAR date at MS-B, July 1970) (BY70)
2	B-1B CMUP-Computer (SAR date at MS-B, May 1996) (BY95)
3	C-5 RERP (SAR date at MS-B, Dec 2001) (BY00)
4	V-22 Osprey (SAR date at MS-B, May 1986) (BY84)
5	CH-47D Chinook (SAR date at MS-B, June 1976) (BY75)
6	ALCM Missile (SAR date at MS-B, Oct 1977) (BY77)
7	OTH-B (SAR at MS-B, June 1982) (BY82)
8	Harpoon Missile (SAR at MS-B, June 1971) (BY70)
9	T-45TS (SAR at MS-B, Oct 1984)(BY1984)
10	YAL-1 (SAR at MS-B, March, 1996) (BY97)
11	SSN 21 Sub (SAR at MS-B, Jan 1989) (BY85)

Appendix H: Preliminary/Main Model with 45 Programs

Response MS-B to IOC (Months) [Regression Output]

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.465725
RSquare Adj	0.412298
Root Mean Square Error	24.84426
Mean of Response	75.13333
Observations (or Sum Wgts)	45

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	21521.715	5380.43	8.7170
Error	40	24689.485	617.24	Prob > F
C. Total	44	46211.200		<.0001*

Parameter Estimates

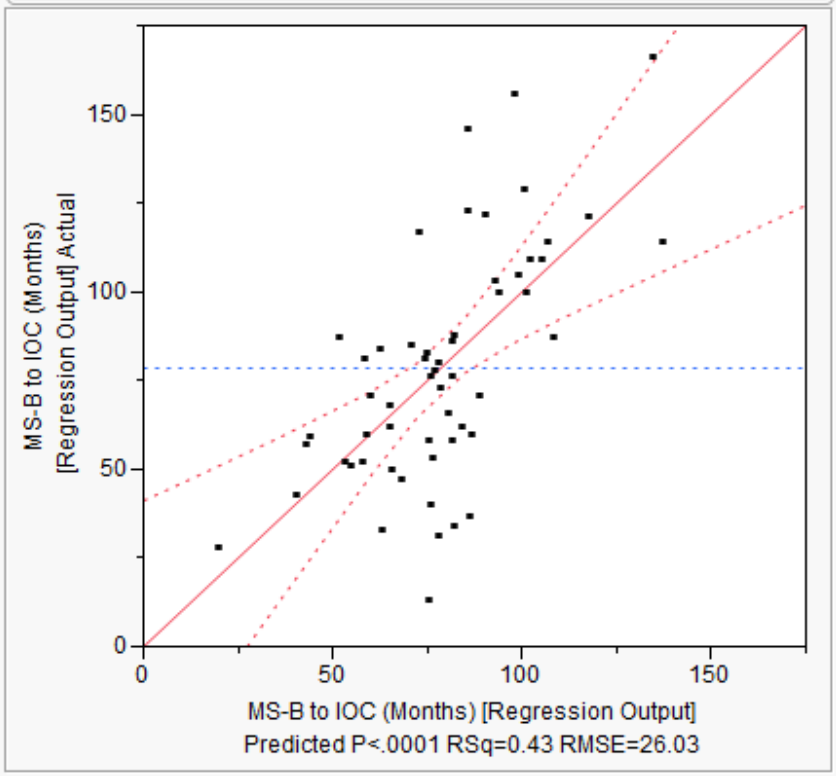
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	82.071139	8.655695	9.48	<.0001*
RDT&E $ (M) at MS-B Start (BY16)	0.0077162	0.002693	2.87	0.0066*
% of RDT&E Funding at MS-B Start (BY16)	-86.70413	24.15955	-3.59	0.0009*
Modification	-19.34527	8.989336	-2.15	0.0375*
1985 or Later for MS-B Start	19.58554	7.635146	2.57	0.0142*

Appendix I: Validated/Final Model with 56 Programs

Response MS-B to IOC (Months) [Regression Output]

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.428807
RSquare Adj	0.384008
Root Mean Square Error	26.03015
Mean of Response	78.48214
Observations (or Sum Wgts)	56

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	25941.974	6485.49	9.5717
Error	51	34556.008	677.57	Prob > F
C. Total	55	60497.982		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	81.938131	7.406529	11.06	<.0001*
RDT&E $ (M) at MS-B Start (BY16)	0.0079896	0.002745	2.91	0.0053*
% of RDT&E Funding at MS-B Start (BY16)	-89.33287	22.93536	-3.89	0.0003*
Modification	-18.34103	8.498096	-2.16	0.0356*
1985 or Later for MS-B Start	24.792533	7.171663	3.46	0.0011*

Bibliography

Bielecki, John V. *Estimating Engineering and Manufacturing Development Cost Risk Using Logistic and Multiple Regression*. MS thesis, AFIT/GAQ/ENC/03-02. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2003.

Birchler, D., M. Christle, and E. Groo, 2011, "Cost Implications of Design/Build Concurrency", *Journal of Cost Analysis*, July, pp. 237-246.

"Bonferroni Correction." *Wolfram MathWorld*. Web. 11 Dec. 2015. <http://mathworld.wolfram.com/BonferroniCorrection.html>.

Brown, Gregory E. *Accuracy of Time Phasing Aircraft Development Using the Continuous Distribution Function*. MS Thesis, AFIT/GCA/ENC/16M. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, 2015.

Brown, Gregory E., Edward D. White, Jonathan D. Ritschel, and Michael J. Seibel. "Time Phasing Aircraft R&D Using the Weibull and Beta Distributions", *Journal of Cost Analysis & Parametrics*, 8:150-164, 2015.

Brown, Thomas W., Edward D. White, and Mark A. Gallagher. "Weibull-Based Forecasting of R&D Program Budgets", *The Journal of Cost Analysis & Management,* 4:1, 41-54, 2002.

Cook, R. D. "Detection of influential observations in linear regression", *Technometrics*, 19, 15–18, 1977.

"Cobra Judy Replacement Achieves Initial Operational Capability." *United States Navy*. Program Executive Office Integrated Warfare Systems, Public Affairs, 1 May 2014. Web. <http://www.navy.mil/submit/display.asp?story_id=80748>.

Cashman, William M. *Why Schedules Slip: Actual Reasons for Schedule Problems Across Large Air Force System Development Efforts*. MS thesis, AFIT/GSM/LAP/95S-2. School of Systems and Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1995.

Defense Acquisition University (DAU)."IOC Definition", *Defense Acquisition University Glossary*. Web. 30 Nov. 2015. <https://dap.dau.mil/glossary/Pages/2048.aspx>.

Deitz, Daniel, Timothy J. Everleigh, Thomas H. Holzer, and Shahryar Sarkani. "Improving Program Success Through Systems Engineering Tools in the Pre-

Milestone B Acquisition Phase," *Defense Acquisition Review Journal*, 3:283-308, 2013.

Drezner, Jeffrey A. and Giles K. Smith. *An Analysis of Weapon System Acquisition Schedules*. Contract MDA903-90-C-0004. Santa Monica CA: The RAND Corporation, December 1990 (R-3937-ACQ).

"Electronic Warfare Products." Lockheed Martin, Electronic Warfare. Web. 11 Dec. 2015 <http://www.lockheedmartin.com/us/products/electronic-warfare.html>.

Fichter, Lewis S, Darryl K. Arnold, and Michael K. Allen *DoD Weapon System Schedule Risk Profiles Based on an Analysis of Schedule Data Presented in the December 1991 Selected Acquisition Reports (SARs)*. Santa Barbara CA: Tecolote Research, Inc., December 1992 (TO-92-007).

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

Genest, Daniel C. *Logistic and Multiple Regression: The Two-step Approach to Estimating Cost Growth.* MS thesis, AFIT/GCA/ENC/04-01. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2004.

Giacomazzi III, Michael P. *An Analysis of the Impact of Defense Acquisition Reforms and External Factors on Schedule Growth of Defense Weapon Systems*. MS Thesis, AFIT/GCA/ENV/07-6M. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, March 2007.

Government Accountability Office (GAO). "Space Acquisitions: DOD Delivering New Generations of Satellites, but Space System Acquisition Challenges Remain." *United States Government Accountability Office*. GAO, 11 May 2011. Web. 04 Nov. 2015. <http://www.gao.gov/products/GAO-11-590T>.

Harmon, Bruce. "The Limits of Competition in Defense Acquisition", pp. 11. *Defense Acquisition University*. Web. 2012. <http://www.dau.mil/research/symposiumdocs/Friar - Cost Growth paper.pdf>.

Harmon, Bruce R., Lisa M. Ward, and Paul R. Palmer Jr. *Assessing Acquisition Schedules for Tactical Aircraft*. Ft. Belvoir VA: Institute for Defense Analyses Paper P-2105, February 1989 (AD-B129975).

Harmon, Bruce R. and Lisa Ward. *Methods for Assigning Acquisition Schedule for Air-Launched Missiles.* Institute for Defense Analyses Paper P-2274, November1989 (AD-B139609).

Harmon, Bruce R. and Neang I. Om. *Schedule Assessment Methods for Surface-Launched Interceptors*. Ft Belvoir VA: Institute for Defense Analyses Paper P-3014, August 1995 (AD-A305991).

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

James, Deborah. "Reinventing the Aerospace Nation." *Secretary of the Air Force*. U.S. Air Force. Web. 15 Sept. 2015. <http://www.af.mil/AboutUs/AirForceSeniorLeaders/SECAF.aspx>.

JMP® Version 10.01, (Academic), CD-ROM. Computer software. SAS Institute Inc., Cary NC, 2011.

Lucas, Brandon M. *Creating Cost Growth Models for the Engineering and Manufacturing Development Phase of Acquisition using Logistic and Multiple Regression*. MS thesis, AFIT/GCA/ENC/04-02. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2004.

McClave, James T., P. George Benson, and Terry Sincich. *Statistics for Business and Economics.* New Jersey: Prentice Hall, 2001.

McDaniel, Chris J. *Estimating Cost Growth in Engineering and Schedule Cost Categories using a Two-pronged Regression Approach*. MS thesis, AFIT/GCA/ENC/04-03. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2004.

Microsoft® Excel 2010. Version 14.0.7163.500 (32-bit). Computer software. Microsoft® Corporation, Redmond WA, 2010.

Monaco, James V. *Predicting Schedule Risk: A Regression Approach.* MS thesis, AFIT/GCA/ENC/05M. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2005.

Monaco, James V. and Edward D. White. "Investigating Schedule Slippage", *Defense Acquisition Review Journal*, 12(2), 176-193, 2005.

Monaco, James V and Edward D. White, III. "Extending Cost Growth Estimation to Predict Schedule Risk," *The Journal of Cost Analysis and Management*, 1:13, 2006.

Moore, Gary W. *Estimating Procurement Cost Growth Using Logistic and Multiple Regression*. MS thesis, AFIT/GAQ/ENC/03-02. Graduate School of Engineering

and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2003.

National Aeronautics and Space Administration (NASA). *NASA Cost Estimating Handbook*. Washington: National Aeronautics and Space Administration, 2002.

Neter, John, Michael H. Kutner, Christopher J. Nachtseim, and William Wasserman, Applied Linear Statistical Models (Fourth Edition), pp. 111-381. Boston: McGraw-Hill Companies, Inc., 1996.

Reig, Raymond W. "A Decade of Success and Failures in the DoD Acquisition System" *Program Manager*, 1: 27-29 (January-February 1995).

Rodrigues, Louis J. *Joint Strike Fighter Acquisition: Development Schedule Should be Changed to Reduce Risks*. GAO/T-NSIAD-00-132. March 2000.

Rossetti, Matthew B. *Logistic and Multiple Regression: A Two-pronged Approach to Accurately Estimate Cost Growth in Major DoD Weapon Systems.* MS thesis, AFIT/GCA/ENC/04-04. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2004.

Sipple, Vincent P. *Estimating Engineering Cost Risk Using Logistic and Multiple Regression*. MS thesis, AFIT/GAQ/ENC/02-02. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2002.

Stine, R. "Graphical Interpretation of Variance Inflation Factors", *The American Statistician,* 49:1, 53-56, 1995.

Tyson, Karen W., Harmon, Bruce R., and Utech, Daniel M. *Understanding Cost and Schedule Growth in Acquisition Programs.* Ft Belvoir VA: Institute for Defense Analyses Paper P-2967, July 1994 (AD-A284321).

Tyson, Karen W., J. Richard Nelson, Neang I. Om, and Paul R. Palmer. *Acquiring Major Systems: Cost and Schedule Trends and Acquisition Initiative Effectiveness.* Contract MDA903-84-C-0031. Alexandria VA: Institute for Defense Analyses, March 1989 (P-2201).

Under Secretary of Defense for Acquisition, Technology and Logistics (USD(AT&L)), *Department of Defense Instruction (DoDI) 5000.02*, January 7, 2015 Edition, page 19, Accessed 1 February 2016: http://www.dtic.mil/whs/directives/corres/pdf/500002p.pdf

US Code: Title 10: Section 2432. 1 Oct 2015

http://assembler.law.cornell.edu/uscode/html/uscode10/usc_sec_10_00002432----000-.html

Unger, Eric J. *Relating Initial Budget to Program Growth with Rayleigh and Weibull Models*. MS Thesis, AFIT/GAQ/EN/01M-3. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB, Ohio, March 2001.

Unger, Eric J., M.A. Gallagher, and E.D. White III, 2004, "R&D Budget-Driven Cost and Schedule Overruns", *Journal of Cost Analysis*, Summer, pp. 80-97.

Vita

Captain Christopher A. Jimenez graduated with a Bachelor of Science (B.S.) degree from the United States Air Force Academy in the Class of 2011. 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 Jimenez has gained a variety of operational experience across the spectrum of financial services, including cost analysis and accounting. 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 Air Force Cost Analysis Agency, Washington, DC.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
21-03-2016	Master's Thesis	Sep 2014 – March 2016

TITLE AND SUBTITLE

Predicting Schedule Duration for Defense Acquisition Programs: Program Initiation to Initial Operational Capability

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Jimenez, Christopher A., Captain, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/ENC)
2950 Hobson Way, Building 641
WPAFB OH 45433-8865

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENC-MS-16-M-161

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Life Cycle Management Center
1865 4th Street, Wright-Patterson AFB, OH 45433
Phone: (937) 656-5478; E-mail: gregory.brown.34@us.af.mil
ATTN: Capt Gregory Brown

10. SPONSOR/MONITOR'S ACRONYM(S)

AFLCMC/AFCAA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRUBTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

Accurately predicting the most realistic schedule for a defense acquisitions program is an extremely difficult task considering the inherent risk and uncertainties present in the early stages of a program. We use a multiple regression analysis to predict schedule duration in a defense acquisition program. The prediction scope of our research is limited to predicting schedule duration from program initiation to initial operation capability (IOC). We use the data from 56 programs across all services, which was acquired from a SAR database created by RAND. We were able to achieve an R^2 of 0.429 and an Adjusted R^2 of 0.384 in our finalized model multiple regression model. Based on the lack of clarity present in the early stages of a program, we look to the R^2 and Adjusted R^2 of our final model to be significant in predicting schedule from program initiation to IOC. We found whether MS-B start occurred in 1985 or later, a program was a modification or not, the % RDT&E funding at MS-B start (BY16), and RDT&E $(M) at MS-B start (BY16) to be significant predictors of the time in months between MS-B start and IOC.

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	104	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