3-2021

An Analysis of the Effects of Technology Readiness Levels on Cost Growth

Christopher R. Bissing

AN ANALYSIS OF EFFECTS OF TECHNOLOGY READINESS LEVELS ON COST GROWTH

THESIS

Christopher R. Bissing, First Lieutenant, USAF

AFIT-ENV-MS-21-M-207

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.

AFIT-ENV-MS-21-M-207

AN ANALYSIS OF EFFECTS OF TECHNOLOGY READINESS LEVELS ON COST GROWTH

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 R. Bissing, BS

First Lieutenant, USAF

March 2021

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

AFIT-ENV-MS-21-M-207

AN ANALYSIS OF EFFECTS OF TECHNOLOGY READINESS LEVELS ON COST GROWTH

THESIS

Christopher R. Bissing, BS

First Lieutenant, USAF

Committee Membership:

Dr. Edward D. White
Chair

Lt Col Scott T. Drylie, Ph.D.
Member

Dr. Jonathan D. Ritschel
Member

Mr. Shawn M. Valentine
Member

AFIT-ENV-MS-21-M-207

Abstract

This research seeks to evaluate the effects of Technology Readiness Levels (TRL) on Cost Growth. It makes use of data from Technology Readiness Assessments (TRA) and Selected Acquisition Reports (SAR) to explore relationships between TRLs at Milestone B and cost growth in Major Defense acquisition Programs (MDAP) and Major Automated Information Systems (MAIS). Programs using higher proportions of critical technologies rated below TRL 7 tend to experience greater cost growth than programs that use more mature technologies. Current DoD doctrine requires TRL 6 to enter Milestone B. The results of this research seek to evaluate the merit of this requirement. TRL usefulness in multiple linear regression models is assessed by comparing against regression models without the use of TRLs. Results indicate relationships between TRL and Cost growth may be driven by omitted variables such as length of EMD phase. A more complete dataset may indicate that TRLs are driving EMD length and provide insight into potential causes of schedule slippage.

Acknowledgments

I would like to thank my beautiful wife, whose loving support help me persevere through the most challenging aspects of this program. I would like to thank my advisor, Dr. Tony White, who provided me with many ideas on how to vector my research and helped strengthen my knowledge in statistics. I would also like to thank my sponsor Mr. Shawn Valentine and Mr. Rob Currie, who provided data sources and suggestions on how to proceed with my research.

Christopher R. Bissing

Table of Contents

List of Figures

Page

List of Tables

Page

List of Equations

AN ANALYSIS OF EFFECTS OF TECHNOLOGY READINESS LEVELS ON COST GROWTH

I. Introduction

Background

Technology Readiness Levels (TRL) have potentially profound impacts on cost and schedule performance for Major Defense Acquisition Programs (MDAP). The Government Accountability Office (GAO) stresses the significance of properly assessing programs for Critical Technology (CT) elements. The GAO provides a case study of a Navy Columbia-class submarine, which failed to identify several technologies as critical. This resulted in the underrepresentation of technical risk. According to the GAO, the development of CTs is key to meeting cost, schedule, and performance requirements (2020).

The concept of TRLs was first conceived by NASA in 1974, and more fully developed in the 1990s (Dunbar, 2017). TRLs measure technology readiness on a scale of 1 to 9 based on technological maturation. (Dunbar, 2012) In addition to assessing the viability of programs, the DoD can use TRLs to predict cost and schedule growth in Major Defense Acquisition Programs (MDAP) and smaller programs. The DoD requires a TRL of at least 6 to enter the Engineering, Manufacturing, and Development (EMD) phase in the Defense Acquisition System, although the Government Accountability Office (GAO) recommends a TRL of at least 7 (GAO, 2020). For this purpose, TRLs lower than 6 are considered technically immature. Higher TRLs are desirable as they pose less of a risk to the government in terms of development costs. More proven technologies also lessen the risk of schedule slippage as they have matured past the conceptual phase. Given the versatility of TRLs, it is worthwhile to study the similarities in programs that experience technological maturation.

Given the relative scarcity of data available on TRLs, there is not a significant body of research on the effect of TRLs on cost and schedule growth. Research Summary Reports (RSR) contain some data on TRLs, but the data is often incomplete or missing. Using TRLs to estimate cost growth is particularly difficult, given there is no uniform cost growth as technology progresses to higher levels. A study found that program cost grows exponentially as a proportion of its development costs as TRLs mature (Linick, 2017). The GAO has published data on TRLs though case studies for various programs, but there is a lack of robust statistical analysis of this data. As technology progresses through TRLs, programs require additional funding to develop and integrate those technologies. For this reason, the DoD seeks more technologically developed solutions in early stages whenever possible to avoid incurring exponential higher costs later in the development process. This is potentially why program managers find modifications to existing systems attractive.

In their 1993 RAND study, Drezner et al. measured performance indices in SARs against cost growth. They hypothesized that lower performing programs would exhibit higher cost growth. The implicit assumption was that performance could be used as a proxy for technical complexity. For poorer performing programs, they inferred the program experienced complications related to technical challenges. Their findings were inconclusive, but they suggest this may have been due to an incomplete measure of technical complexity. Our research builds on the RAND study by using TRLs as a proxy for technical complexity. We hypothesize lower TRLs are associated with higher technical complexity due to the inclusion of novel technologies. Higher technical complexity should lead to higher cost growth. If TRLs do in fact help explain technical complexity, they would only represent a small component of this complexity. Additional tools to consider might be Manufacturing Readiness Levels (MRL), and Systems Readiness Levels (SRL), but those are beyond the scope of this research.

Problem Statement

Given that initial TRLs vary across programs, it would be beneficial to analyze TRLs at program initiation to examine how they might contribute to cost growth. Specifically, an analysis of TRL data could yield information on the types of contracts, programs, program length, and program acquisition phases that are most associated with high technology maturation. This information could directly lead to better cost and schedule estimates by incorporating TRLs in risk analysis.

Research Objective

The main purpose of this research is to build on existing predictive cost growth models by incorporating TRLs into analysis. Our research will begin by building a multiple regression model using explanatory variables for cost growth. We will then seek to determine if adding TRL variables to the model improves explanatory power. We should note that our objective is not build a model to predict cost growth at any level. We simply seek to gain a basic understanding of how TRLs may add to our knowledge of cost growth predictors. This is a rudimentary analysis and more research will be needed before we can build predictive models.

Research Questions

To determine how TRLs at program initiation can potentially drive cost growth, the following research question is considered:

1. How do initial TRLs affect cost growth in Major Defense Acquisition Programs and Major Automated Information Systems in the EMD phase?
2. What TRL predictor variables are the more significant cost or schedule drivers?

Methodology

The data on TRL is primarily provided by the Air Force Research Laboratory (AFRL) and AFLCMC. The use of SARs and Technology Readiness Assessments (TRA) will be essential in documenting and assessing TRLs. The analysis will include descriptive statistics of TRL and other quantitative variables. We will begin by building a regression model using proven cost predictors such as total acquisition costs, weapon type, and program quantities. After building an initial model, we will incorporate TRL variables to analyze how predictive power in the model has grown. We will do this by evaluating adjusted R^2 values. Because we are not building a model for predictive purposes, we will not reserve a portion of our data as a test set. The model will serve as a foundation for further research. Additionally, we will evaluate categorical TRL variables against cost growth variables. This will help answer our second research question, explain which TRL components have the most significant effects on cost.

Assumptions and Limitations

Data for TRLs is obtained through Technology Readiness Assessments and data provided by AFLMC. The extent to which this research produces meaningful results will largely depend on the completeness and availability of data. Difficulties in past research have stemmed from having incomplete information. Appropriate steps will have to be taken to account for missing data. The spurious nature of some quantitative variables warrants a careful interpretation of results. Conclusions must be derived from a logical application of theory. For example, TRL increase may drive cost growth, but it would seem erroneous to conclude that cost growth necessitates technology maturation.

Organization

This thesis seeks to determine how TRLs can provide an explanation for cost growth. Understanding these characteristics may aid in decision support for program selection and cost estimation. This chapter provided a brief overview of TRLs and identified the TRL thresholds for programs entering the EMD phase. It identified gaps in existing TRL research and explained historical limitations on TRL analysis. The research questions listed above will provide a framework with which to assess the data. Chapter II provides an overview on current research of TRLs, existing TRL models, and will introduce how this research will improve on those models. Chapter III provides data sources, descriptive statistics on data, and details the methodology used to analyze the data. Chapter IV offers an interpretation of the results of data analysis, and Chapter V will summarize and conclude this research and provide answers to the research questions above.

II. Literature Review

Technology Readiness Levels and Technology Readiness Assessments

Technology maturity is a fundamental consideration in government acquisitions. Technological progress can have considerable implications to program cost and ultimate success. Because there are fewer unknowns with more advanced technologies, program managers and decision-makers take less risk by incorporating these technologies. The DoD and other government organizations have independently developed metrics to evaluate technology readiness. One of the most well-known metrics is NASA's Technology Readiness Levels (TRL). TRLs are a systematic metric that provide a standard for assessment of technology maturity (Mankins, 1995). Although the concept of TRLs was first introduced by NASA in the 1970s, NASA further defined TRLs in the 1990s on a scale from 1 to 9 (Appendix A). At TRL 1, only basic principles of technology are observed and reported. At TRL 9, the technology has been flight proven in successful mission operations (Mankins, 1995). In the early 2000s, the DoD started using TRLs in their technology assessments. The adoption of a standardized system of measurement not only helps to evaluate technology consistently, but provides an objective means to compare between materiel solutions. TRLs signal information about cost and schedule risk.

TRL is the basis for the Government Accountability Office's (GAO) Technology Readiness Assessment (TRA). The GAO defines TRA as an evidence-based process that evaluates the maturity of technologies critical to the performance of a larger system or the fulfillment of the key objectives of an acquisition program, including cost and schedule.

(2020) The TRA assigns TRLs at different milestones in a program, most importantly to support program initiation. TRAs cannot eliminate technology risks, but they can help identify potential areas of concern as programs move forward. Using TRAs to mitigate risks can save the government a substantial amount of time and money as it gets more costly to take corrective action as a program progresses.

Before we discuss prior research on TRLs, we have to provide a background on cost growth. Cost growth has been a subject of study for RAND, IDA, and many other organization for the past several decades. It is of special interest to program managers and policymakers, as inaccurate cost estimates can lead to overfunding or underfunding programs. This does not mean inaccurate cost estimates are the sole source of cost growth, but it is one culprit of interest (McNicol, 2004). The following research summarizes just a handful of the many studies conducted on cost growth.

Cost Growth

Light, Leonard, Pollak, Smith, and Wallace (2017)

Light et al.'s research is particularly relevant to our study as they use multivariate regression to quantify cost and schedule risk. The four variables they analyze are whether the program completed a Milestone A review, the shared of the planned development budget expended prior to MS B, the share of development budget planned concurrently with production, and whether the program is joint or single-service. Contrary to prior research (Drezner, 1993), they actually found that ongoing programs had higher cost and schedule growth factors than completed ones.

Light defines concurrency as the percentage of RDTE funding years in which Procurement is also funded. There are proponents of concurrency who believe that by simultaneously engaging in development and production, costs can be reduced by compressing schedules. The opposing argument predicts concurrency leads to cost growth as a result of re-work and out-of-sequence manufacturing. In their study, they found a negative relationship between concurrency and cost growth when they included ongoing programs. They caution that this does suggest program managers should incorporate concurrency into program schedules in an effort to mitigate cost growth. Programs that plan for concurrency may be inherently less risky and are more adaptable to design changes even after production has started.

They found a 10% point increase in RDTE budget allocated prior to Milestone B corresponds to a 4.5% decrease in Program Acquisition Unit Cost (PAUC) growth. They also find that programs that have a higher share of their budget dedicated to RDT&E experienced less cost growth, but more schedule growth. Despite their findings, they note the limitations of using their regression model as a predictive tool, as the model contained a large amount of unexplained variation.

Jimenez (2016)

Jimenez also used multiple linear regression in his thesis on pre-milestone A predictors for schedule growth. Although his research focuses on schedule, and not cost growth, his methodology and conclusions provide useful insights for our research. Jimenez build s a regression model to predict schedule from MS B to IOC using pre-Milestone A data. His model, constructed using stepwise regression, indicated RDTE funding up to MS B, %RDTE Funding at MS B, and modification were significant variables. Another significant variable was a dummy variable indicating if the program had started after 1985. This is mean to account for temporal effects, specifically the policies of the 1985 Blue Ribbon Commission. In our research we will also include a temporal variable for 2009, indicating if the effects of the Weapons System Acquisition and Reform Act has a significant impact on cost growth.

Jimenez finds a positive coefficient for RDTE $M at MS B, and hypothesizes this may have something to do with technology maturity of the program. Programs that are not technologically mature may incur schedule delays. More initial RDTE funding may be an attempt to mitigate longer schedules. The opposite effect is found when considering the percentage of RDTE that is front-loaded before MS-B. Of the four variables. %RDTE Funding at MS B has the strongest influence. A greater percentage of RDTE allocated at Milestone B corresponds to a shorter duration from MS B to IOC. This translates to a greater investment in TMRR, which aims to reduce technology, engineering, integration, and life-cycle risk. Jimenez theorizes an increased investment proportionally toward TMRR puts a greater emphasis on increasing technology maturity. Additionally, modification programs had a negative correlation with schedule. This makes intuitive sense, as modification programs are more likely to start with operationally-tested technology.

Although he did not find any significant findings for concurrency, it's possible this may be a predictive factor for cost growth. Jimenez also notes he would like to have known the TRL of a program at MS B, and it was the most significant variable not available in SARs. Our study fills this gap by using TRL data from TRAs. He concludes that program managers could use this multiple regression model to predict what program schedules should be, look for potential efficiencies, and provide a cross-check for on-going programs.

Kozlak (2016)

Kozlak calculated cost growth factors from the Development Estimate to several other key milestones throughout aircraft programs. He first makes use of logistic regression to analyze which programs sustained cost growth, then uses contingency tables to analyze which categorical variables had explanatory power. Among the variables Kozlak considered were: % RDTE appropriation funded at Milestone B, the time in months estimated from MS B to IOC, the procurement quantity at MS B, and the RDTE, Procurement, and Total Estimates at Milestone B. He focuses on data measurable at Milestone B to ensure his results are applicable to policy.

Kozlak plots CGF against % program completion at different points throughout the program. As programs matured through the development and production phases, procurement cost growth was higher than the expected value as the program approached IOC, and stabilized as it neared the finish. Although Kozlak does not comment directly on this, this relative cost decrease could be attributed to learning that occurs in production after IOC. Development cost growth followed a different pattern, at times growing at a slower rate than program completion.

Using logistic regression, Kozlak found different predictive variables of cost growth depending on the point in time in the program that cost growth factor was evaluated. Recurring predictive variables were %RDTE Funding at MS B, Procurement Qty/Months (Program Length), Procurement Qty at Milestone B, and months from Milestone A to Milestone B. To further analyze these variables, Kozlak created binary categorical variables assessing each variable at certain break points. These break points were determined by evaluating descriptive statistics for each variable. In addition to using variables that were significant with logistic regression, Kozlak includes other categorical variables to analyze with contingency tables. The variables include aircraft type, prototyping effort, whether MS B started before 1985, service (Air Force), and modification. Results varied dependent on funding categories, but he found that variables were more predictive of development growth than procurement or total cost growth.

Kozlak concludes that median % total cost growth at IOC is 47%, 114%, and 91%, for procurement, development, and total funding categories, respectively. This suggests that cost growth is much greater for procurement than development for aircraft programs. He also notes IOC has a median program completion percentage of 48%. While our study does not analyze purely aircraft data, we can use this study to compare how cost growth is incurred at the various programs' midpoints compared to aircraft.

DeNeve, Ryan, Ritschel, and Kabban (2015)

The major critique DeNeve et al. have of traditional cost growth studies is that most studies assume baselines will not change, even though this is a common occurrence. They employ a unique risk-estimating methodology through macro-stochastic modeling. This seeks to increase estimate accuracy by creating models for several different groups. Distinct models are created based on varying degrees of cost growth factors, with the smaller cost growth programs using one model and the largest cost growth programs using another model. In this way, they are able to improve accuracy of their model by tailoring to programs with similar cost growth characteristics. The parameters they include are service, development to production ratio, acquisition cost, quantity change, and year count. The year count parameter ensures the model is robust to trends across time.

They categorized programs by commodity type, modification, number of years funded, and joint-service. By employing their model, they were able to reduce estimate error by 37% at Milestone B. Because more information is known about the program as it nears completion, the models lose power over time. At the end of the program, their models reduce estimate error by 19%. Because they grouped programs by increasing degrees of cost growth, the model showed the most significant results for high cost-growth programs. DeNeve et al recognize their model is not suitable for low-level cost estimates, but it can reduce cost estimating error in early phases of the program.

Cancian (2010)

Cancian comments on cost growth literature to identify the importance and root causes of cost growth. He suggests cost growth may not reflect poor management control, but rather poor initial estimates. These poor estimates do not indict the cost estimator, as there are many programs costs that are hard to predict. Cancian mentions the research and development efforts for the F-22, which saw a 50% increase in RDTE funds alone. Although these changes are difficult to predict, TRLs may provide another component that help predict accurate cost and risk assessments.

Cancian suggests delaying production until development is complete to prevent retrofitting. He also suggests that accurate estimates are important because they provide critical information in an Analysis of Alternatives. Cost growth acts a tax on programs, causing them to reduce quantities, reduce testing or cut down on support equipment. Cancian claims Technology maturity is an acquisition strategy that helps manage programs better (2010). TRLs are especially useful because managers have access to this information before a program begins.

Younossi, Arena, Leonard et. al (2007)

Younossi et al aim to answer the questions: 1) What is the cost growth of DoD Weapon systems? And 2) What has been the trend of cost growth over the past three decades? Their study focuses on Air Force programs, measuring completed as well as ongoing programs. To ensure meaningful comparisons, they measure cost growth from MS B to 5 years past MS B for all programs. In alignment with past research practice, they normalized for inflation and quantity using learning curves. The found the programs experienced a rapidly increasing portion of their development and procurement cost growth in the first 40% of the program, as indicated in Figure 1 below. This relates to our study, as we aim to measure cost growth only during EMD.

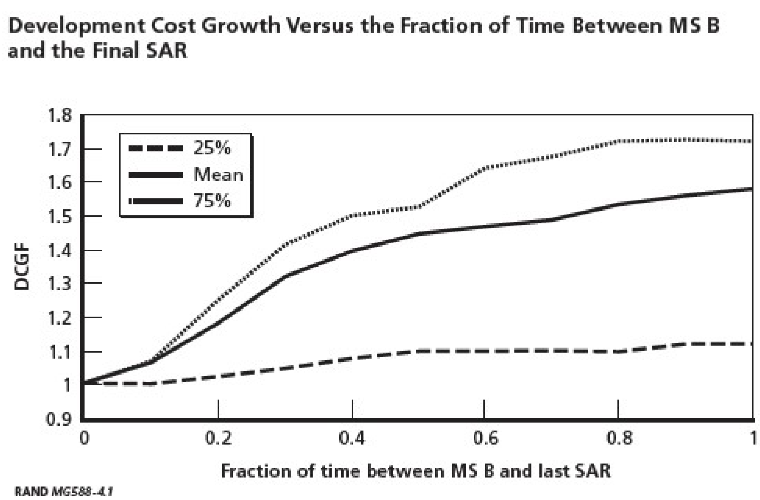


Figure 1: Development Cost Growth (Younossi)

They noted that the mix of programs by commodity changed over time. The 1990s had more electronics programs, which past studies have shown have realized less cost growth. When looking at commodity, they found ongoing space and aircraft programs experienced higher development cost growths than the average completed programs (including all commodities). On the other hand, missile programs experienced lower than average cost growth.

They found "on average 60 percent of cost growth occurred when the program was about one-third of the way between MS B and the last SAR". They used a 5-year past MS B at a measuring point for CGF because most programs in their data were beyond the early design stage and preliminary design review at that point. They also note from the completed programs that most cost growth occurs early in the development phase. They concluded in their analysis of completed programs that longer programs experienced higher than average development cost growth. They also concluded electronics programs had the lowest cost growth. They find that because cost growth had remained high from the 1970s through the 1990s, acquisition reform policies failed to offer much improvement.

Arena, Leonard, Murray, and Younossi (2006)

Arena analyzed 68 completed programs from a RAND SAR database for cost growth from Milestone II and Milestone IIIa to the end of the program. These milestones closely mirror current Milestones B and C. His results indicate higher cost growth in development versus procurement appropriations. Cost growth factor means and standard deviations were smaller from MS III than MS II, indicating higher accuracy and precision.

Arena was interested in analyzing cost growth improvement as programs moved from Milestone II to the final SAR. Because of the varying lengths of time between Milestone II and the last SAR, he divided this time period into 11 equidistant segments from 0% to 100. The 0% segment evaluated cost growth from Milestone II to the last

SAR, while the 50% segment evaluated cost growth from the median between Milestone II and the last SAR. As one might expect, cost growth factors decreased as the approached the end of the program because the baselines are being gradually pushed up. At around 70 to 80% program completion, the CGF averages 1.0 and observes the same CGF as the final program cost. Therefore, they find at around 80% program completion, continued cost growth is negligible.

When measuring for correlation, Arena took the LN(CGF) to account for the lognormal distribution of CGFs. In the RDTE appropriation, program duration, MS Year, and commodity had significant correlation with MS II. Electronics programs tended have lower cost growth, while satellites and launch vehicles experienced higher cost growth. Longer programs likely exhibit higher cost growth as they are subject to more revisions, upgrades, and potential technology obsolescence. Programs with more recent Milestone IIs had lower CGFs. Arena cautions against misinterpreting these findings, as more recent programs in the 1990s are necessarily shorter than previous decades. Once again, electronics programs proved to have lower cost growth from MS II. Although prior studies found a relationship between program size and cost growth, Arena did not find these results in his study.

Average total cost growth for completed programs was 46% from Milestone II and 16% from Milestone III. The sample in this study demonstrated higher cost growth than Drezner's study, providing support for the importance of analyzing complete programs.

Drezner (1993)

Drezner analyzed SAR data to assess general cost growth trends. He found a strong association between program size and cost growth, using total acquisition cost as a barometer for program size. He notes smaller programs tend to have higher cost growth. He suggests this may be because smaller programs receive less management oversight than larger dollar value programs. For every year past EMD start, program cost increased by about 2.2% per year. He notes this is relevant because a distorted CGF is calculated if program age is not accounted for. To control for this, Drezner's study includes only programs that are at least 3 years past EMD start. Programs that are younger than this will not have experienced significant cost growth due to a short amount of time to experience potential problems. Drezner finds CGFs of 1.35, 1.2, and 1.16 for Army, AF, and Navy programs, respectively. Additionally, the pooled average is 1.2. This suggests Army programs on average realize greater cost growth compared to the other services. Drezner posits this is likely attributed to smaller and older programs in the Army. He then breaks down cost growth by commodity, noting cost growth for vehicles is considerably higher than the average cost growth across all commodities. Again, Drezner ties this back to service, by explaining vehicle programs are dominated by the Army and are typically smaller and older programs.

Comparing 5 programs at different baselines, Drezner finds cost growth to be 1.40, 1.32, and 1.09 for Planning, Development, and Production estimates, respectively. This seems to confirm that cost estimation, at least in those programs is systematically biased downward, but again he concedes that these programs are also typically older and have smaller dollar values. Additionally, modification programs have the benefit of using data from existing programs to inform the estimate. He found modification programs on average have slightly lower CGFs that new programs.

Longer programs allow more time for unanticipated events to occur that affect cost performance. Program managers may try to mitigate cost growth by planning for concurrency in their programs, an overlap between the EMD and Production phases. The implications of this strategy are unclear, as a shorter program may help dampen cost growth, but starting production without a finished development phase may lead to technical problems. Comparing between programs that did and did not have concurrency, he found no significant difference, however among the subset of programs that did have concurrency, a higher overlap resulted in lower cost growth. He measures concurrency as overlap between the completion of IOT&E and Milestone 3a start date. Drezner plots schedule slip against cost growth and finds no correlation. This implies schedule slippage and cost growth may be independent of each other, and the factors that affect one do not necessarily affect the other.

When measuring programs with one prime contractor versus programs with multiple contractors, jointly-managed programs saw less cost growth, contrary to the expectation. RDTE dollars tend to experience greater cost growth than procurement dollars because the technical problems are worked out during the development phase. Program size and maturity are the two factors most strongly associated with cost growth after inflation and quantity are adjusted. Drezner states that because program size and maturity and dominate other factors affecting cost growth, they must be considered in analysis.

Hough (1992)

In his 1992 study, Hough lists various limitations associated with using SAR data. He goes in depth into each one, but for they are listed below. Our study would not be complete without acknowledging these limitations. Arena (2006) summarizes Hough's limitations below.

1. **Aggregate Data**
2. **Baseline Changes, Modifications, and Restructuring**
3. **Reporting Guidelines, and Requirement Changes**
4. **Inconsistent Allocation of Cost Variances**
5. **Incomplete or Partial Weapons Systems Cost**
6. **Exclusion of Certain Types of Programs**
7. **Ambiguity of the Estimate Bias**
8. **Unidentified Risk Reserve**

In addition to mentioning the drawback of SARs, Hough also mentions the importance of consistent baselines. Hough references the Bradley Fighting Vehicle System (BFVS) as an evolutionary outgrowth of the Mechanized Infantry Combat Vehicle (MICV). Instead of creating a new SAR for the Bradley, the programs office reported cost data for the Bradley on the original MICV. Given the largely different system requirements and capabilities, the cost growth of the BFVS looked excessively large when using the MICV DE as a baseline. In this case, a new Development Estimate was justified for the BFVS, as it could be argued it was an entirely new system. The BFVS was not consistent with the way the MICV was originally defined. Unlike this clear example, it is not always obvious when evolutionary changes justify using a newer DE as a baseline estimate. For the purposes of our study we, choose to use the DE at the original MS B. The only exception is if the program had its milestone B re-structured due to any number of reasons. One reason programs need Milestone B re-approval is for Nunn-McCurdy breaches, like the F-35. In this case, the original Milestone B is not indicative of the current state of the programs and using the re-approved Milestone B provided a more accurate standard of comparison.

TRL Studies

There is far less research on TRLs that cost growth, but to give a brief introduction, some a few studies relevant to our research are summarized below.

Katz (2013)

Katz et al. study the relationship between technology and design maturity on cost growth and schedule slippage in DoD weapons systems. This was done primarily with the goal of providing guidance for planning and execution during the EMD phase. Katz et al define Relative Schedule Change as the percentage difference between the final schedule length and the initial schedule estimate. They used the GAO's guidance and indicated low-maturity if the TRL was less than 7, and high maturity if the TRL was 7 or greater. Although their findings proved inconclusive for TRLs as a driver for cost change, they found TRLs did have an influence on schedule slippage. Relative schedule slippage from Milestone B to Milestone C yielded p-values of .05 and .03 when measuring TRL at Milestone B and the CDR, respectively. Importantly, there is a disparity between what the DoD and GAO consider as technologically mature. The DoD accepts a TRL of 6 as sufficiently mature, while the GAO requires a TRL of 7. Katz study only finds a statistical significance in schedule slippage between mature and immature technologies when using the GAOs definition of mature technology. This provides further evidence for the importance of program managers using a universal metric when assessing technologies. What is interesting to note is that when looking at the timeframe from MS B to MS C, the number of mature technologies increases from 16 to 28 from the start of MS B to the CDR. This points to technology maturation during the EMD phase. Katz does not directly address this in his research, but understanding the components that led to maturation could help explain the significance of schedule slippage. Katz et al note their research is limited to TRL change, cost growth and schedule slippage during the EMD phase.

Dubos et. al (2008)

Dubos at al aimed to study the correlation between TRL and schedule slippage in space systems. They then identify appropriate schedule margins commensurate with TRL risk. Low TRLs are commonly associated with schedule risk. Increased technology maturity leads to lower schedule risk due to increased knowledge concerning technology development and manufacturing. This provides higher confidence that projects will be completed on schedule (Dubos, 2008). Schedule runs can often occur with low-TRL programs because it is difficult to predict how long it will take to progress from the low-TRL research phase to the high-TLR development phase. Dubos et al. use initial TRL, initial schedule estimate, and final duration to model schedule slippage as a function of TRL. They determined that higher accuracy of initial schedules was correlated with higher TRLs and therefore had lower mean schedule slippage. Dubos believes the TRLs have omitted variable bias, and there are other factors that may be driving schedule slippage. He notes that identifying these factors would be relevant as they would help limit schedule risk and identify best practices to maintaining acquisition programs on schedule. Dubos et al. defined schedule risk as the probability that the actual schedule would exceed a point estimate plus a schedule margin. Given their results, they were able to develop schedule-risk curves that could predict an appropriate schedule margin to include if they wanted the schedule risk to remain at a certain level for a given TRL.

Smoker and Smith (2007)

Smoker and Smith created a difference-in-differences table to quantify the amount of time if took the GPS program to advance from one TRL to the next. They then regressed the log of Estimate Development Cost, log of Estimated Procurement Cost, and log of Estimated Total Program Cost against time. They hypothesized an exponential relationship between time and final system cost. Given that the cost of maturing technologies grows exponentially over time, they suggest programs with higher initial TRLs will experience lower cost growth.

They also sought to determine if different program types required different amounts of time to progress from one TLR to the next. They analyzed three solid rocket programs, a launch vehicle program, two sensor satellite programs, a comm satellite program and GPS User Equipment to determine cost-growth factors. The results of their analysis revealed there were substantial differences in time required to mature technologies. Their research has practical applications for those who want to use TRL cost-growth factors to incorporate technology maturation into cost estimates. Previously, cost estimates were based solely on the current knowledge of technological advancement. Smoker and Smith's data provides a method to account for technology maturation. One could directly apply their results by using cost-growth rates from analogous programs to estimate cost-growth for new programs.

GAO Technology Readiness Assessment Guide

Critical Technologies

In its 2020 TRA Guide, the Government Accountability Office (GAO) illustrates the importance of properly identifying CTs. The proper identification of CTs is a crucial starting point to mitigate risk as programs move through development. The report states, “Not identifying the technologies as critical means Congress may not have had the full picture of the technology risks and their potential effect on cost, schedule, and performance goals as increasing financial commitments were made” (GAO, 2020).

The most common mechanism TRA teams use to identify CTs is a Work Breakdown Structure (WBS) (GAO, 2020). Lower-level WBSs can be used to identify technologies within a system. The GAO lists several benefits from using a WBS including availability, comprehensiveness, and showcasing the relationships between the parts, end product, and tasks. There is some degree of subjectivity in determining CTs, as the criteria is specific to individual organizations. The GAO developed a list of questions from multiple government TRA guides to use as a baseline in defining CTs. If the answer is “yes” to at least one question in the following two lists, the program should consider the element under review in its initial list of CTs.

Technical questions:

1. Is this technology new (for example, next generation)?
2. Is the technology used in a novel way?

3. Has the technology been modified?
4. Is the technology expected to perform beyond its original design intention or demonstrated capability?
5. Is the technology being used in a particular or different system architecture or operational environment than it was originally intended or designed for?
6. Could the technology have potential adverse interactions with systems with which it will interface?

Programmatic Questions:

1. Do requirements definitions for this technology contain uncertainties?
2. Does the technology directly affect a functional requirement?
3. Could limitations in understanding this technology significantly affect cost (for example, overruns) or affordability?
4. Could limitations in understanding this technology significantly affect schedule (for example, not ready for insertion when required)?
5. Could limitations in understanding this technology significantly affect performance?

This list demonstrates a potential framework that could be applied consistently across organizations to standardize the definition of CTs. Currently, the GAO TRA Guide is a guideline and there is no requirement to follow these practices. Although this standardization would help normalize the CT identification process, there is still some subjectivity in definitions. For example, different services may have different ideas of whether a technology is used in a novel way. This leads to a realistic possibility of an inconsistent evaluation of TRLs across different programs.

Summary

This chapter reviewed the most relevant cost growth and TRL literature in relation to our research. Far more information was available on cost growth, which has been studied for decades. Some cost growth studies, such as Cancian, Drezner, and Jimenez, expressed the need to evaluate TRLs and their effect on cost growth. Chapter III will explain how this study plans to build on the limited body of knowledge on TRLs.

III. Methodology

The purpose of this chapter is to describe the data sources and techniques used to build the models for analysis. The first part will detail the data collection and normalization process. The second portion describes the methods used to assess the data. The two main sources of data for this effort were the Acquisition Information Repository (AIR), and Selected Acquisition Reports (SAR) obtained from the Defense Acquisition Management Information Retrieval (DAMIR) system. The AIR contains a variety of cost and technical data for MDAPs. AIR was the primary source of TRL data for this research. In addition to these sources, the Air Force Life Cycle Management Center (LCMC) provided a consolidated SAR database used for cost and numerous other variables. This is an internal database compiled by members of AFLCMC. The AFLCMC also provided over 480 Program Office Estimates (POE) used to expand our TRL data. The majority of POE data did not provide additional value, as it was mostly ACAT II and III programs. We focused on data for ACAT I Programs to make use of SARs. Although limited, the POE data did provide information on a few select ACAT I programs. The main method of analysis was regression using Ordinary Least Squares (OLS) methodology. We also created contingency tables by measuring categorical variables against different breakpoints is our response variables. To reiterate, our goal is not to create a model for predictive purposes. This is a fundamental study seeking to determine if TRL data can help predict cost and schedule growth.

Of the 32 programs we analyzed, only 6 were complete through 90% of their funding. However, only 7 programs had not yet completed Milestone C, so we felt confining our study to EMD was the best way to provide a consistent standard of comparison. For the purposes of this research, EMD is defined as the period between MS B and MS C, although we acknowledge Milestone C definition varies by service and program. Many programs mark the completion of IOT&E as the end of EMD, but not all programs have this date indicated in their SARs. Most programs in our data had a Milestone C date. Measuring from MS B to MS C offers some consistency, despite its flaws.

TRL Data

TRL data was obtained primarily through the AIR database. To access AIR, users must first login through the Defense Acquisition Visibility Environment (DAVE). DAVE is an extensive database that provides access to information for acquisition reporting, analysis, insight, and decision-making. It serves as a centralized source to access various resources such as AIR, DAMIR, and the Cost Assessment Data Enterprise (CADE). The AIR is a valuable source for milestone acquisition information. It provides access to milestone certifications and recommendations, Nunn-McCurdy breaches, cost estimates, and Technology Readiness Assessments (TRA) among other documents. The AIR contained records for 91 TRAs. Many of these TRAs are conducted by Independent Review Teams and were primarily used to inform Milestone B decisions. As such, there is no standardized format for TRAs. Some records were lengthy reports detailing TRA methodology, while others were memorandums that simply documented TRLs for CTs. One challenge that became quickly apparent in the data collection process was an inconsistency of how independent review teams rated CTs. As mentioned previously, there is no standard procedure for evaluating critical technologies. Some programs may consider certain technologies critical, while others will not.

Because the primary response variable of interest is cost growth in the EMD phase, we looked for TRAs performed at milestone B. Of the 82 records pulled, 67 contained TRL data. Some of these records only confirmed that a TRA was conducted but provided no further information. Other programs reported their systems did not contain technology that met the definition of critical technology elements set forth in the Technology Readiness Assessment Guidebook.[1] We also excluded reports that were not conducted to support a Milestone B decision. Some reports recorded TRAs at Milestone C or Critical Design Reviews (CDRs). While some TRAs were clearly labeled as a Milestone B supporting document, others required further investigation. Fortunately, most TRA or TRA reviews documented the date the original TRA was conducted. We matched this date with the Milestone B date of the program to confirm we were using the data from the appropriate timeframe. The dataset contained a few duplicate records that provided no additional value. It is worth mentioning that ten of these records were TRAs for the same programs at two different points in time. This would allow the calculation of potential change in TRL from one phase to another. Unfortunately, there were insufficient pairs of programs like these to make for useful analysis. Of the remaining 33 programs, 2 did not have cost data available from DAMIR or AFLCMC's SAR database. We eliminated the DDG-1000 from our dataset, as it is a canceled program. Finally, we

[1] The 2020 TRA Guide provides the most current definition for a critical technology: Critical technologies are technology elements deemed as critical if they are new or novel, or used in a new or novel way, and are needed for a system to meet its operational performance requirements within defined cost and schedule parameters. These technology elements may be hardware, software, a process, or a combination thereof that are vital to the performance of a larger system or the fulfillment of the key objectives of an acquisition program. (GAO, 2020)

were able to obtain TRL data for 2 additional programs using the AFLCMC POEs. In the end this left 32 programs available for analysis. Table 1 summarizes the exclusion criteria. The final programs contained SARs ranging from 1996 to 2019 A full list of programs is in Appendix B.

Table 1: Program Exclusion Criteria

Criteria	ΔTRA Reports	Total TRA Reports
Total Reports Pulled	82	82
Contained No TRL Data	-15	67
TRA Not at Milestone B	-16	51
Duplicate Record	-18	33
No Cost Data Available	-2	31
Canceled Program	-1	30
POE Data	2	32
Final Reports (n)		32

SAR Data

SARs provided a convenient metric to assess growth across different programs. SAR data were obtained through DAMIR and an AFLCMC cost database. DAMIR contains SAR data broken down into full SARs, Acquisition Program Baselines (APB), SAR Baselines, and other metrics. This was useful in creating response variables and even some independent variables. The LCMC database includes data from 4477 SARs from 1970 to 2019. The original sources for these data were DAMIR and RAND. The SAR database included information to help normalize data such as inflation indices, quantities, baselines, ACAT, and commodity. This information helped generate a set of control variables for regression analysis.

There were some drawbacks associated with the SARs in our sample. Most programs had SARs available on DAMIR, but a few SARs were accessed through DAVE's SAR/MAR catalog for historic SARs. The SARs in DAVE were generally for completed programs. These SARs were not as conveniently broken down as DAMIR's database, which allowed us to filter specifically for APB and Baseline estimates. To access most recent APB values in DAVE, we looked at the APB Development Estimates in the most recent SAR available. Many of our key predictive variables required funding data. While SARs for current programs contained this data, many SARs for completed programs did not. This limited the usefulness of using independent variables that required funding information.

TRL Assumptions

Some TRAs did not contain TRLs for every CT in the system, but that did not prevent those TRAs from being useful. For example, the Navy's P-8A had a TRA which cited 27 TRLs, only ten of which had not met TRL 6. Nine of those ten had a documented TRLs, but one CT had a TRL labeled as "N/A". A TRL average could still be calculated by using 26 CTs instead of 27, as we were missing information for one CT. Additionally, the other 17 CTs did not contain a TRL rating. The TRA only mentioned that these technologies were at least TRL 6 and demonstrated minimal risk since they had been used previously in other programs. For this reason, we assign a TRL of 6 to all technologies labeled as technologically mature but otherwise not rated. This method has potentially drawbacks, as those demonstrated technologies may have higher TRLs that could skew our data.

For software programs in particular, it can be difficult to perform a TRA for programs ahead of a Milestone B decision. For the NSA's Key Management Infrastructure, the independent review team stated it was unable to make an assessment on the TRL for three CTs. This is because the team was not entirely sure what technologies were going to be incorporated until after a contract was awarded. The review team conducted their analysis using analogous technologies from similar programs. For our data, we assessed these CTs to be level 6 based on the review team noting that these technologies had been used in the past.

For various reasons, we had to make assumptions on TRLs for some programs. The availability of information is scarce as it is, and we could not afford to throw any data out. Until we have access to more detailed data, this may limit the application of our results.

Response Variables

The main response variable of interest in this study is EMD cost growth. We measured this is several ways. Mimicking the methodology of several RAND and IDA studies, we broke cost growth into RDTE, Procurement, and Total Estimates. Some programs have negligible Milcon and Acquisition O&M accounts that we incorporated into total estimates, but we did not measure these individually. For our regression, we are mainly interested in total cost growth. We use the individual appropriations to determine any trends. Cost was measured as a cost growth factor- the program estimate at Milestone C divided by the estimate at Milestone B. Additionally, we use the natural log of cost growth factors to account for the fact that cost growth is usually lognormally distributed with a right skew (Arena et al., 2006). All dollars are normalized to BY 2021.

TRL Variables

Programs do not generally receive an overall TRL rating. Rather, individual CTs within programs receive ratings. This leaves options for how to annotate TRLs. One way to measure TRL is to choose the lowest TRL as a limiting factor. Alternatively, you could note the highest TRL as a potential cost driver. A measurement approach that offers the most variability in the data is to take an average TRL. Figure 2 shows a distribution for the aforementioned TRL variables.

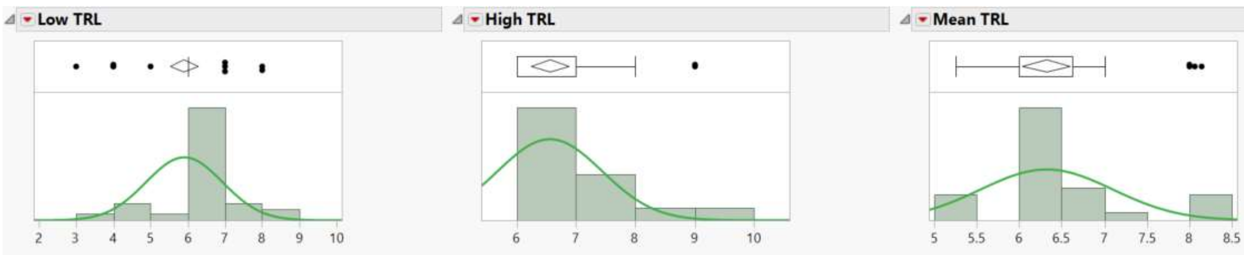


Figure 2: Histograms for Lowest, High, and Average TRL

In addition, individual programs were assessed by the percentage composition of each TRL in the program. For example, a program with a CT rated TRL 6 and another CT rated TRL 7 would be 50% TRL 6 and 50% TRL 7. Table 2 describes the TRL variables initially considered for analysis.

Table 2: Description of TRL Variables

Variable	Type	Description
Low TRL	Continuous	Lowest TRL in Program
High TRL	Continuous	Highest TRL in Program
Average TRL	Continuous	Average TRL in Program
% TRL Below 6	Continuous	% TRL Below 6 in Program
% TRL 6	Continuous	% TRL 6 in Program
% TRL 7	Continuous	% TRL 7 in Program
% TRL 8	Continuous	% TRL 8 in Program
% TRL 9	Continuous	% TRL 9 in Program
% TRL <=6	Continuous	% TRL <=6 in Program
Low TRL <7	Categorical	1 if Low TRL <7
High TRL<7	Categorical	1 if High TRL <7
Mean TRL <6.3	Categorical	1 if Mean TRL <6.3 in Program

Other Independent Variables

Evaluating the effect of TRL variables alone is insufficient to conclude model validity. Additional variables are assessed to account for possible omitted variable bias. By comparing models with and without TRL variables, we can determine if adding TRL provides any additional benefit. If not, it is possible TRL effects are captured by other variables. A series of independent variables helped serve as predictors and controls for cost growth. The variables were selected based on previously tested relationships, as described in Chapter II. The predictive variables are those that are known at Milestone B and can be used to predict cost and schedule growth. The control variables were either not measurable until later on in the program, or they were outside of the program manager's control (such as Service). These include EMD Sched Actual, % Program Complete by Funding, and EMD Slippage. Control variables are necessary as they help explain other potential causes of cost growth. EMD Length, for example, might suggest that the time from Milestone B to Milestone C may be the primary driver of cost growth as opposed to one of the predictor variables. The independent variables are listed and described below.

- *RDTE ($M) Funding MS B – Continuous Variable (Predictive)*
  - Continuous variable indicating the raw amount of RDTE funding allocated before Milestone B.
- *%RDTE Funding Milestone B – Continuous Variable (Predictive)*
  - Continuous variable indicating the % of the total Program's RDTE funds allocated through Milestone B (as of the Milestone B estimate)
- *RDTE % of Total Budget MS B – Continuous Variable (Predictive)*
  - Continuous Variable indicating the percentage of the Total Milestone B Estimate that was appropriated to RDTE
- *% Years Funded MS B – Continuous Variable (Predictive)*
  - Continuous variable representing the percentage of funded program years through Milestone B divided by the total number of planned funding years
- *MS Start Year – Continuous Variable (Predictive)*
  - Continuous variable indicating the year of the Program's most recently approved Milestone B
- *MS B > 2009? – Binary Variable (Predictive)*
  - Binary Variable, "1" if MS Start Year is after 2009
- *Est EMD Length Continuous Variable (Predictive)*
  - Continuous variable indicating the predicted length of EMD as of Milestone B
- *Production Quantity MS B – Continuous Variable (Predictive)*
  - Continuous variable indicating the planned Production quantity
- *Proc MS B – Continuous Variable (Predictive)*
  - Continuous variable indicating the procurement estimate at Milestone B
- *RDTE MS B – Continuous Variable (Predictive)*
  - Continuous variable indicating the RDTE estimate at Milestone B

- *Total Estimate – Continuous Variable (Predictive)*
  - Continuous variable indicating the total acquisition estimate at Milestone B

- *MS A? – Binary Variable (Predictive)*
  - This variable indicates if a Milestone A was formally documented in the SAR. A value of "0" indicates no Milestone A was documented but does not confirm no Milestone A review was performed.

- *Commodity – Binary Variables (Predictive)*
  - A dummy variable was assigned to the Aircraft, Ship, MAIS (Electronic), Vehicle, Satellite, Missile, Munition, and Helicopter commodities. The "Other" commodity category is omitted. A value of "1" indicates the program was the corresponding commodity.

- Concurrency
  - A continuous variable indicating the percentage of the RDTE Funding Years where Procurement is also funded. For example, if RDTE Funding is from 2011 to 2020, and Procurement is from 2016 to 2030, Concurrency is 50% because 5 out of 10 of RDTE's funding years (2016 to 2020) also fund Procurement.

- *Contractor – Binary Variables (Control)*
  - A dummy variable was assigned to Boeing, Northrop Grumman, Lockheed Martin, General Electric, and Raytheon. The omitted category was "Other". Note that we primarily use contractor in our exploratory analysis, as there is nothing that leads us to believe one contractor is inherently more costly than another, all else equal.

- *Joint Contract – Binary Variable (Predictive)*
  - A dummy variable indicating if the program had more than one major contractor

- *Service – Binary Variables (Control)*
  - A dummy variable was assigned to Army, Navy, and DoD. "Air Force" was the excluded service. A value of "1" indicates the program was the corresponding commodity. For joint-service programs, the lead service is used for classification.

- *MS C Complete? – Binary Variable (Control)*
  - A dummy variable with a value of "1" if Milestone C had been completed. Programs with a value of "0" used data available from the latest SAR, the most recent point in the program.

- *% Program Complete by Funding – Continuous Variable (Control)*
    - Continuous variable indicating the percentage of the program that has been funded at Milestone C, or the most recent point in the program if the program has not reached Milestone C.
- *EMD % of Program Length – Continuous Variable (Control)*
    - Continuous variable indicating the length of EMD relative to the entire program. Because most program in this data are not yet completed, this metric is an approximation based on most recently available data.

The main control variables we included were EMD Sched Actual, Milestone C Complete?, and % Program Completed by Funding. Because we include programs that have not completed Milestone C, we include this variable to determine if EMD completion makes a significant difference in cost and schedule growth during the EMD phase. The % Program Completed by Funding allows us to determine if the observed growth is associated with program maturity through Milestone C.

Table 3 shows the distribution of commodity types across different services, and Figure 3 shows a breakout of programs by commodity type. Half of all programs were either Aircraft or electronic. Seven of the eight programs in the electronic category were Major Automated Information Systems (MAIS). The dataset also includes variables that account for time components. Including these variables helped adjust for different EMD lengths or time between the first and last SAR. Many variables were removed throughout the model selection process, as they revealed multicollinearity and reduced degrees of freedom. Only four programs experienced growth in production quantities. For those that did, we normalized by subtracting the variance associated with quantity from the procurement and total estimates at MS C. We chose to normalize to the Development Estimate as opposed to the Production estimate to maintain a consistent baseline

Table 3: Distribution by Commodity

Distribution by Commodity					
Commodity	Air Force	Army	Navy	DoD	Total
Aircraft	4	0	3	1	8
Electronic	3	0	2	3	8
Helicopter	0	0	1	0	1
Missile	0	1	0	0	1
Munition	1	0	0	0	1
Satellite	2	0	0	0	2
Ship	0	0	4	0	4
Vehicle	0	2	0	0	2
Other	2	2	1	0	5
TOTAL	12	5	11	4	32

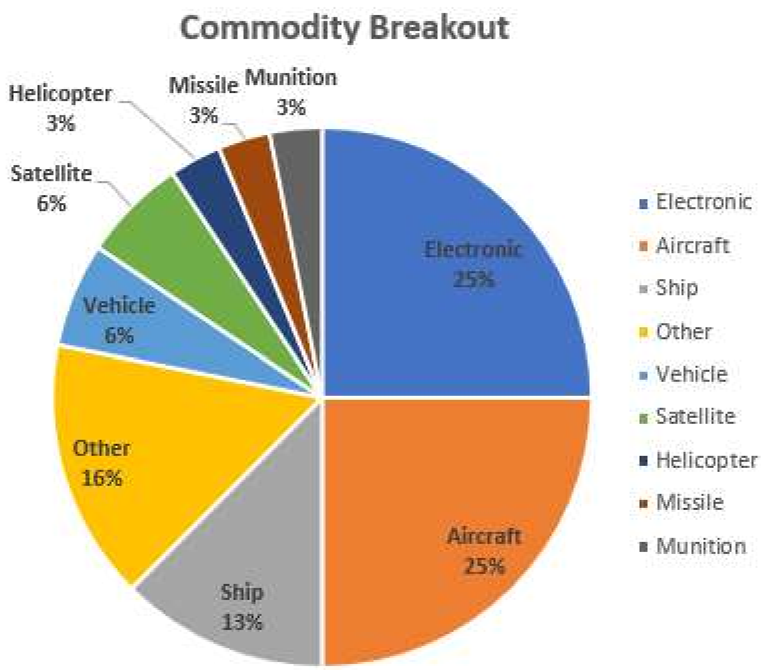


Figure 3: Commodity Breakout

Analysis Methods

Before performing stepwise regression, we ran individual bivariate analyses to get am initial impression of significant relationships. We tested TRL variables and the other independent variables against LN (RDTE CGF), LN (PROD CGF) and LN (Total CGF). We also tested against schedule length and schedule slippage. This resulted in over 500 individual regressions. The results of this analysis are in Appendix C. Testing for these relationships first made it easier to build as stepwise models, as we knew which variables to include. This was especially important given we had over 50 different independent variables and only 32 data points. Our small dataset limits our degrees of freedom available.

Regression Analysis

The main method of analysis we focus on is OLS Regression. We start by inputting the significant variables from our previous analysis into JMP's Fit Model tool, and perform mixed stepwise regression to introduce only significant variables into the model. After we run an initial test, we analyze VIF Scores, correlation matrices, Cook's Distance, and significance values to eliminate more variables one at a time. For the purposes of this study, we use a α=.1 value as our criterion to keep variables in the model. We choose a larger benchmark than the standard α=.05, because too stringent criteria gives little power to detect effects in small samples (Light, 2017).

We conduct this process for a model with and without TRL variables. This allows us to compare the two and assess which model has the most explanatory power. The general econometric specification for the model is

$$\hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \cdots + \hat{\beta}_k x_{k,i} \tag{1}$$

This model is an estimate for the true population parameters, which also includes an error term, ε:

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i} + \varepsilon \tag{2}$$

Our goal is to try to determine the best values for these parameters, while confirming OLS assumptions to ensure our results are not biased.

Contingency Table Analysis

Contingency tables allow us to test the relationships between two categorical variables. In our research, we aim to test categorical TRL variables against cost growth variables. We discretized TRL and LN(Total CGF) into distinct categories to allow for this two-way analysis. This allowed us to isolate specific subsets of our data that might be driving cost growth. The mosaic graph allows us to visually depict the difference in means for each category. Fisher's exact test allows us to test for significance in small sample sizes.

Summary

We discussed the sources and types of data used to build our models. After filtering the data for usability, we removed TRAs that were not viable for analysis. Ultimately, we have 32 programs to analyze for a variety of TRL and control variables. Cost growth is our primary response variable and is denoted by LN(RDTE CGF), LN(Prod CGF), and LN(Total CGF) variables. Because the data were imperfect, we had to make some assumptions to account for incomplete information. Analysis will be conducted primarily through multiple linear regression and contingency tables. Chapter IV will describe the variables in more detail and display the results of our analysis.

IV. Analysis and Results

Overview

This chapter presents the results of the statistical analysis described in Chapter III. First, descriptive statistics provide an overarching view of the data. We use bivariate analysis to get initial impressions for independent and TRL variables against LN(Total CGF). Next, the results of the multiple linear regression analysis provide some insight into the predictors of EMD cost growth. We finish our regression by validating OLS assumptions. We then explain the statistically significant contingency tables. To conclude the chapter, limitations to the data will be evaluated as well as possible solutions for improved analyses.

Descriptive Statistics

The results of these descriptive statistics are limited by the small sample size, but we tabulate them in Table 4 anyway to provide perspective. One notable observation is a Low TRL minimum of 3. Normally a TRL 6 is the minimum acceptable standard for entry into Milestone B. Occasionally, MDAs will make exceptions for TRLs below 6 when there is a technology maturation plan in place to achieve TRL 6 by CDR. Using the natural log transformation helps account for the right-skew in cost growth.

Table 4: Descriptive Statistics

Variable	N	Min	Mean	Median	Max	Std Dev
Low TRL	32	3	5.906	6	8	1.027
High TRL	32	6	6.563	6	9	0.876
Mean TRL	32	5.25	6.321	6	8.143	0.778
% TRL Below 6	32	0	0.051	0	0.5	0.137
% TRL 6	32	0	0.674	0.917	1	0.391
% TRL 7	32	0	0.169	0	1	0.297
% TRL 8	32	0	0.08	0	1	0.261
% TRL 9	32	0	0.026	0	0.5714	0.109
% TRL <=6	32	0	0.717	1	1	0.414
% TRL >6	32	0	0.283	0	1	0.414
LN(RDTE CGF)	32	-0.2236	0.07	-0.002	0.723	0.221
LN(Proc CGF)	32	-1.19	-0.04	0	0.579	0.319
LN(Total CGF)	32	-0.572	0.037	0	0.607	0.233

Before we conducting bivariate analyses, we summarize trends for cost growth factors and schedule factors by TRL to give a perspective on how TRLs affect these measures. The CGFs and schedule factors in the tables below provide mean values for the subset of data analyzed. For example, the mean total CGF for the 4 programs with a mean

TRL below 6 is 1.108. Generally, higher mean TRLs are associated with decreasing CGFs. Curiously, total CGF is high for mean TRLs above 8, but given that this represents only 2 programs, these could be anomalies.

Table 5: CGF Summary Statistics by TRL

CGF Summary Statistics by TRL				
Variable	**n**	**RDTE CGF**	**Proc CGF**	**Total CGF**
Mean TRL Below 6	4	1.087	1.071	1.108
Mean TRL = 6	16	1.092	0.977	1.027
6 < Mean TRL <=7	8	1.137	0.926	1.07
7 < Mean TRL <=8	2	0.991	1.005	1.004
Mean TRL > 8	2	1.161	1.388	1.343
Low TRL < 7	27	1.108	0.97	1.048
Low TRL >= 7	5	1.06	1.184	1.166
High TRL <= 7	20	1.091	0.996	1.043
High TRL > 7	12	1.117	1.016	1.104

The most significant trends are those associated with schedule factors. As seen in Table 6, increasing TRLs correspond to decreasing EMD lengths. Average EMD length drops 53% from 5.27 years for programs with a Mean TRL below 6, to 2.47 years for programs with a mean TRL above 8. Programs with lowest TRL at least equal to 7 have an average of 68% of the EMD length of programs with lowest TRLs less than 7. Programs with high TRLs at least equal to 7 have an average of 73% of the EMD length as programs with High TRLs below 7.

Table 6: Schedule Summary Statistics by TRL

Schedule Summary Statistics by TRL				
Variable	**n**	**EMD Est (Years)**	**EMD Actual (Years)**	**EMD Slippage (Factor)**
Mean TRL Below 6	4	5.19	5.27	1.01
Mean TRL = 6	16	3.7	4.59	1.206
6 < Mean TRL <=7	8	3.43	3.77	1.109
7 < Mean TRL <=8	2	3.54	3.17	0.976
Mean TRL > 8	2	2.33	2.47	1.024
Low TRL < 7	27	3.81	4.47	1.161
Low TRL >= 7	5	3.28	3.05	0.975
High TRL < 7	20	4	4.73	1.167
High TRL >= 7	12	3.26	3.44	1.073

Bivariate Analysis

Before creating the multiple linear regression models, we performed exploratory analysis to evaluate which variables could be candidate cost drivers. This gave us a foundation for variables to include in the regression models. The results of this analysis are summarized in Appendix C. Quite surprisingly, none of the TRL variables were significant at even α=.1 for LN(Proc CGF), LN(RDTE CGF), or LN (Total CGF). Among the other independent variables, variable significance varied by the response variable. The LN(RDTE), for example, had significant associations with Est EMD Length, EMD Length Actual, and EMD slippage. The LN(Total CGF) was also significant for Est EMD Length and EMD Actual at α=.1, but not for slippage. Several commodities such as Ship, Satellite, and Helicopter had recurring significant relationships against the many response variables. The usefulness of this information is questionable as many commodities like Missiles and Helicopters only had 1 unit in the data. The results gave us enough information to start populating our stepwise regression models. Notably, whether or not MS C was complete, or the % of Program complete by funding were not significant factors for LN(Total CGF), although they seemed to have an effect on RDTE and Proc individually.

Stepwise Regression

Given the multitude of variables to choose from, performing stepwise regression helped pare our model down to leave the significant variables. We used our data to generate as many predictive variables as possible, as justified by prior research. Rather than including every possible variable into the model and running the regression, we preliminarily eliminated variables that would cause multicollinearity issues. Before running stepwise regressions, we analyzed correlation matrices to flag any high correlation between independent variables. Correlation above .7 was worth a closer look. After confirming correlation would not cause substantial multicollinearity (some will always be present), we input our variables in JMP's stepwise regression tool.

To evaluate the effect of including TRL variables in a regression model, we first built a model without TRL variables as a baseline. We compare the explanatory power of the models by looking at the Adjusted R-Squared values. After running the models using the stepwise regression approach, we analyzed the variables left in the model. First, we looked at Cook's Distance to evaluate if any data had a disproportionate influence on the results. Our initial model had 2 variables that highly influential data points as indicated by Figure 4. The analysis of variance for the first model is shown in Figure 5.

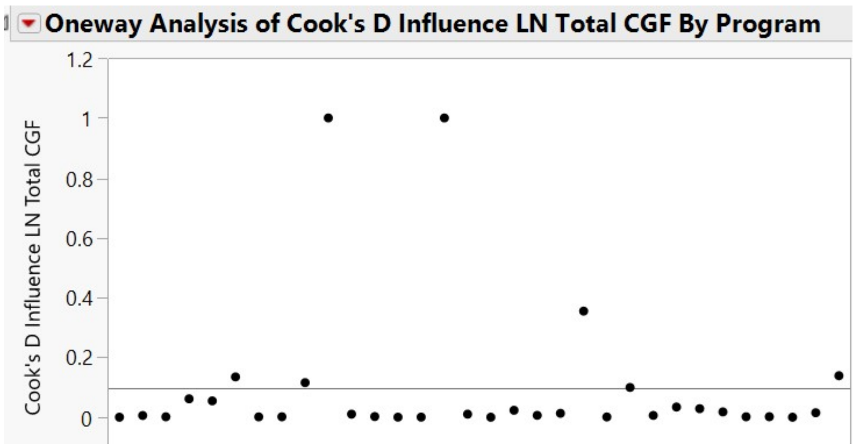


Figure 4: Cook's D Values

At this point, we removed the Enhance Polar System and OCX programs from the model at their Cook's D value of 1.0 were far above the rule-of-thumb threshold of 0.5. Both programs' Cook's D values were more than three times the standard deviation of values across all programs. After adjusting the model for these programs, we came to our preliminary model in Figure 5. Note the Variance Inflation Factors are all below 2, which is well within the threshold for a valid model.

Response LN Total CGF

Effect Summary

Lack Of Fit

Summary of Fit

RSquare	0.300406
RSquare Adj	0.219683
Root Mean Square Error	0.190268
Mean of Response	0.018355
Observations (or Sum Wgts)	30

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.4041742	0.134725	3.7215
Error	26	0.9412546	0.036202	Prob > F
C. Total	29	1.3454288		0.0238*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.1201485	0.047957	2.51	0.0188*	.
DV_Ship	-0.299102	0.106518	-2.81	0.0093*	1.0864928
DV_MAIS	-0.157094	0.082614	-1.90	0.0684	1.106017
Proc Qty at MS B	-6.919e-6	3.529e-6	-1.96	0.0607	1.0571107

Figure 5: Non-TRL Model

Validating the Model

Our model needs to pass OLS assumptions to maintain validity. Note the Variance Inflation Factors are all below 2, which is well within the threshold for a valid model. For an OLS model to be valid, the data must be independent of each other, the distribution of residuals should be normal and should have constant variance. We can test for these assumptions using the Shapiro-Wilks test and Breusch-Pagan tests, for normality and constant variance, respectively. Additionally, the studentized residuals below show no cause for concern for outliers, as all residuals fall within 3 standard deviations of the mean, as shown in Figure 6.

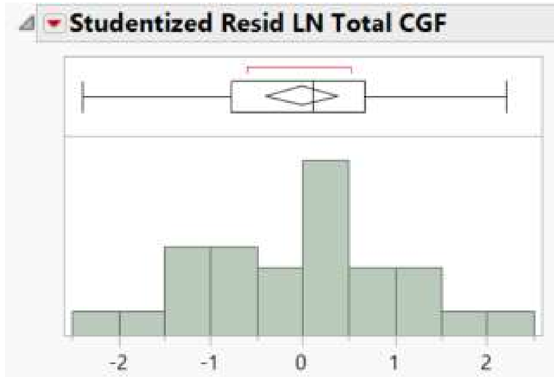


Figure 6: Studentized Residuals

Figure 7 shows the results of the Shapiro-Wilk test. The null hypothesis is that the data is normally distributed. Because we have high p-values, we fail to reject the null hypothesis.

Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	μ	-0.013278	-0.403411	0.3768559
Dispersion	σ	1.0256428	0.8139291	1.3871324

Measure	
-2*LogLikelihood	82.766968
AICc	87.228507
BIC	89.50156

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.991343	0.9967

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Figure 7: Shapiro-Wilk Test

Lastly, although JMP does not have a built-in Breusch-Pagan test for homoskedasticity, we run this test manually using excel. We run the same regression using the sum of the squared residuals as the response variable. We then use the R-Square value for this new regression to calculate a Chi-Square test statistic. Using this test statistic and our degrees of freedom, we obtain a p-value, as seen in Table 7. The null hypothesis of the Breusch-Pagan test is homoskedasticity. We fail to reject this assumption and conclude heteroskedasticity is not present in our model.

Table 7: Breusch-Pagan Test

Breusch-Pagan test	
n	29
Degrees of Freedom	3
R Square	0.136532061
Chi-Square test statistic	3.95943
P-Value	0.256878

TRL Model

We repeat this process but include TRL variables the second time. We did not expect many significant TRL variables given the results of our exploratory analysis. Any TRL variables that are significant should be a result of interaction with other variables. Running a mixed stepwise regression using all variables used to create the first model plus High TRL, Mean TRL, and TRL <=6 proved inconclusive. Initially, our TRL model did prove to be significant for %TRL<=6, but we re-ran the model after determining it was influenced by JLTV which had a Cook's D of 1.79. The modified model was only significant for Ship and Joint Contract. This is not surprising, as our initial analysis did not determine any TRL variables were significant in explaining cost growth in EMD. Unfortunately, we are unable to use regression analysis to gain insight on how TRLs may affect cost growth. Unfortunately, we are unable to use regression analysis to gain insight on how TRLs affect cost growth. It seems erroneous to conclude TRLs cannot explain cost growth, as these results are likely a product of our limited data.

Contingency Tables

Another tool we apply for analysis is the use of contingency tables. Contingency tables allow us to evaluate the effect of one categorical variable on another. By looking at histograms of TRL variables and our response variables, we were able to determine logical "breakpoints" for analysis, such as mean and median. The results of our exploratory bivariate analysis suggest there are not many significant relationships between TRL and cost growth, at least given our data and period of analysis. Table 8 summarizes the TRL and response variables used for this analysis.

Table 8: CGF Variables Considered

TRL Variables	Response Variables
74% TRL <=6	RDTE CGF > 1
100% TRL <=6	RDTE CGF > 1.08
Low TRL <7	Proc CGF > 1
High TRL <7	Total CGF > 1
Mean TRL < 6.3	Total CGF > 1.04
	PAUC CGF > 1

None of the 30 combinations of contingency tables produces significant results. This is unsurprising given our bivariate analyses did not predict any significant relationships. Although results proved inconclusive for cost growth, we did find that TRL variables had an impact on schedule variables when we excluded the F-35. Excluding the F-35 was necessary as it has been in system development since 2001, far above the EMD lengths of all other programs. Table 9 shows the variables we considered for Contingency Table Analysis when measuring TRLs against EMD schedule and slippage. We chose to measure when EMD was greater than 3.5 and 4.25 years because those are the median and mean values, respectively. Likewise, the average % TRL <=6 was 74%. This is because most programs were classified as TRL 6.

Table 9: EMD Contingency Variables

TRL Variables	Response Variables
74% TRL <=6	EMD Actual > 3.5
100% TRL <=6	EMD Actual > 4.25
Low TRL <7	EMD Slippage > 1
High TRL <7	
Mean TRL < 6.3	

Our analysis yielded 3 significant contingency tables. There was a significant relationship between %TRL <=6, High TRL<7, Low TRL<7 and the response EMD Actual at α=.05. The results of this analysis are summarized in Table 10. Significant Contingency Tables mosaics can be found in Appendix D. One example is Mean TRL < 6.3 plotted against EMD Actual > 4.25 in Figure 8. To summarize, whenever mean TRL was not less than 6.3, the program never experienced more than 4.25 years (the mean) EMD length. If mean TRL was above 6.3, programs experienced above mean EMD schedule roughly half the time.

Table 10: Contingency Tables Results

* = 2-tailed significance R = Right-tailed significance	EMD Actual > 3.5	EMD Actual > 4.25	EMD Slippage > 1
74% TRL <=6			
100% TRL <=6		R	
Low TRL <7			
High TRL <7		R	
Mean TRL < 6.3		*R	

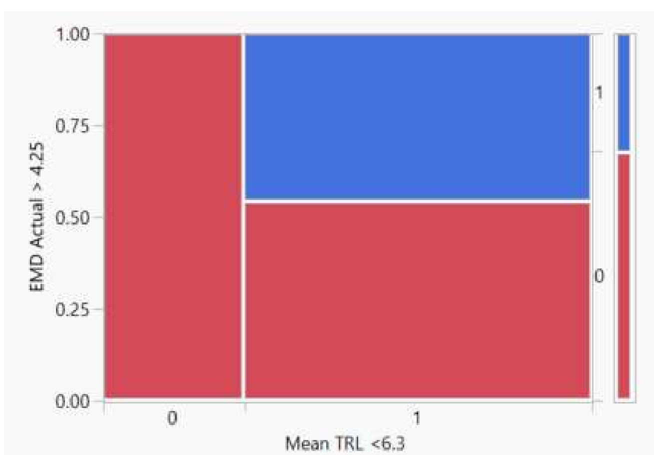


Figure 8: Mean TRL vs EMD Actual

Summary

This chapter provided descriptive statistics and multiple regression and well as contingency table analysis results. We immediately realized our results would be limited by analyzing the bivariate regressions. Surprisingly, TRLs did not have much of any explanatory power at all, but we do not believe this indicates TRLs cannot explain cost growth. We use descriptive statistics to describe general trends of TRLs on cost growth and schedule. Generally, as the TRLs increase, cost growth and schedules decrease. TRLs did not have any significant effect on schedule slippage. Although results were inconclusive for cost growth, we found TRL variables had a correlation with actual EMD schedule length. Programs with mean TRL > 6.3 never experienced higher than average EMD length. Incorporating higher TRLs into system design is associated with shorter program schedules.

V. Conclusions

This chapter presents the results of our analysis in the context of our research questions. Although our TRL data did not have any correlations with cost growth, we did find that they had an association with EMD length. We discuss these results further in this concluding chapter. We document the challenges encountered throughout this process and provide recommendations for future research.

Research Question One

Our first research question sought to identify the relationship between TRLs and cost growth. Our stepwise regression model built with TRL variables did not provide any significant results with regard to cost growth. This is not entirely unexpected, as we have addressed the concerns with having sufficient variability and accuracy of TRL data. We also note that our study is confined to looking at the EMD phase. Much of program cost growth continues to persist after production has started. Therefore, we would expect to see more significant results if we measured cost growth throughout completed programs. It seems unreasonable to conclude that TRL does not play a role in explaining cost growth. With a more robust dataset, we expect future studies will find a negative association between higher TRLs and cost growth.

Research Question Two

Our second research question further defined our task by asking which specific TRL variables predicted cost growth. As we have mentioned, TRLs did were not predictive of cost growth in this study, but they did relate to EMD length. Using programs exclusively comprised of TRLs of 6 or below led to higher EMD lengths than those programs that had at least one CT that was above 6 at program initiation. Similarly, programs with average TRLs higher than 6.3 at Milestone B never experienced higher than average cost growth across all programs. This may be because some programs with lower TRLs experience abnormally long EMD schedules and are skewing the data, so more research is required in this area. Still, our preliminary results suggest it's not a good idea to proceed to Milestone B if the program is not technologically mature, even if there is technology maturation plan in place. These findings support the statutory requirements to achieve TRL 6 for program approval.

Limitations

The process for determining the best relationship using TRL data was not without its limitations. The most quickly apparent limiting factor was lack of availability of TRL data. While US Code Title 10 requires MDAPs to have sufficiently mature technology to enter EMD, there is no regulation mandating formal documentation of TRLs. Although TRAs are conducted and briefed to the MDA, few records are publicly available, leaving us with a small dataset. Along these same lines, there is inadequate variability in the TRL data. One of the OLS assumptions is that data cannot be perfectly multicollinear (Hilmer, 2014). For this assumption to hold, there must be variability in the values of each independent variable. While this assumption is not technically violated because there is minor variability in the data, it does not provide a very robust variable for analysis. We remedied this by calculating an average TRL for each MDAP, but as many critical technologies were simply denoted as TRL 6, many of these average values ended up being TRL 6 as well. We posit that these CTs are often labeled as 6 because this is the threshold required for EMD entry. Stakeholders may not care to further scrutinize the exact level of technological maturity once they have met the base requirement, but this is merely a suggestion worthy of further investigation.

Additionally, some TRAs provided a TRL for each independent CT, or at least gave a breakout of how many CTs were assessed a respective TRL. The Joint High-Speed Vehicle, for example, indicated 12 CTs with TRL 6 and 5 CTs with TRL 7. This made it convenient to calculate average TRL and a % for each TRL. Other programs did not provide the same level of detail and were not as useful for analysis. Lastly, although we initially intended to measure a TRL change from MS B to MS C, many programs do not have TRAs for MS C. Some program offices simply state that a TRA was already conducted at MS B and therefore a new TRA was not required. This made it impossible to assess a TRL change, but perhaps this could still be analyzed with better access to data in the future.

Another problem arises when you consider how the teams that performed TRAs defined Critical Technologies. Although the GAO does provide recommendations as outlined in Chapter II, there is no universally applied definition of a critical technology which leaves it open to interpretation. Different programs or services may differ in how they define CTs. Independent review teams often documented their programs did not contain technologies that met the TRA Guidebook's definition of CTs. This is problematic because it could mean that these programs contained technology elements that are in fact TRL 8 or 9, but they were not deliberately itemized because they posed lesser risk.

There were also some inconsistencies in how different programs delineated milestones. One of our variables of interest was the date of MS C, as we could use this information to form an estimate for EMD length. A handful of programs did not have a MS C, but used another metric to mark the end of the EMD process. Satellite programs, like the Enhanced Polar System, usually did not contain a milestone C. For the EPS, we used the "DT&E Completion for Single String" date as the culmination of EMD. In other programs, the MDA waived the MS C requirement, which required either interpolating a would-be MS date or using another milestone like Full-Rate Production in its place. A list of programs with a conjectured date for Milestone C is annotated in Appendix B.

Another limitation to our analysis is changing baselines. Whenever possible, we strived to use the development and production estimates consistent with Milestone B and C dates. Often, development baselines change throughout the course of the program, but we used the original baseline estimate as our standard. The only exception was when a program restructure forced the program to re-certify its Milestone B at a later date. Because programs restructure often occurs as a result of significant programs changes, we always use the most recent Milestone B date. This has the effect of masking schedule slippage, as a newly established Milestone B will not make it appear as if the program has incurred as much schedule growth.

Future Research

Given the limited availability of TRL data, our program data includes many ongoing programs. It would be preferable to study completed programs to gain an accurate estimate of final actual costs. Prior studies have found that including ongoing programs biases cost growth downward (Arena et al., 2006). A larger dataset would provide a more robust base for analysis. Future research should seek to estimate correlations between TRLs at Milestone B and final cost growth and program lengths for completed programs. It should also look for correlations between TRLs and schedule slippage for the length of the program. Even though our regression model did not produce significant TRL results, it is clear there would have been limited applications in its applicability. We only had 32 programs and about 25% of those programs were software systems. In this research, the software programs experienced less cost growth than the average of all programs, but software programs often have large variances in cost growth. This data is likely not representative of the actual composition of MDAPs. When future studies analyze TRLs they should attempt to create unique models for different commodities, if possible. Lastly, given that we found TRLs are correlated with EMD length, future studies could attempt to build a multiple regression model for EMD length instead of cost growth. Using a similar methodology to that used in this study, you could use TRL variables to determine if TRLs provide additional explanatory power for EMD length.

Appendix A – TRL Definitions

Level	Definition	DoD DAG Description
1	Basic principles observed and reported	Lowest level of technology readiness. Scientific research begins to be translated into applied research and development. Examples might include paper studies of a technology's basic properties.
2	Technology concept and/or application formulated.	Invention begins. Once basic principles are observed, practical applications can be invented. Applications are speculative and there may be no proof or detailed analysis to support the assumptions. Examples are limited to analytic studies.
3	Analytical and experimental critical function and/or characteristic proof of concept.	Active research and development is initiated. This includes analytical studies and laboratory studies to physically validate analytical predictions of separate elements of the technology. Examples include components that are not yet integrated or representative.
4	Component and/or breadboard validation in laboratory environment.	Basic technological components are integrated to establish that they will work together. This is relatively "low fidelity" compared to the eventual system. Examples include integration of "ad hoc" hardware in the laboratory.
5	Component and/or breadboard validation in relevant environment.	Fidelity of breadboard technology increases significantly. The basic technological components are integrated with reasonably realistic supporting elements so it can be tested in a simulated environment.
6	System/subsystem model or prototype demonstration in a relevant environment.	Representative model or prototype system, which is well beyond that of TRL 5, is tested in a relevant environment. Represents a major step up in a technology's demonstrated readiness.
7	System prototype demonstration in an operational environment.	Prototype near, or at, planned operational system. Represents a major step up from TRL 6, requiring demonstration of an actual system prototype in an operational environment such as an aircraft, vehicle, or space.
8	Actual system completed and qualified through test and demonstration.	Technology has been proven to work in its final form and under expected conditions. In almost all cases, this TRL represents the end of true system development. Examples include developmental test and evaluation of the system in its intended weapon system to determine if it meets design specifications.
9	Actual system proven through successful mission operations.	Actual application of the technology in its final form and under mission conditions, such as those encountered in operational test and evaluation. Examples include using the system under operational mission conditions.

Appendix B – MDAP and MAIS in Dataset

1. Improved Turbine Engine Program
2. Joint Precision Approach Landing System
3. F-15 Eagles Passive/Active Warning and Survivability System
4. Advanced Pilot Trainer
5. Common Infrared Countermeasure
6. Armored Multi-Purpose Vehicle
7. Enhanced Polar System*
8. F-22 Modernization Inc 3.2B
9. Ship to Shore Connector
10. Joint Lightweight Tactical Vehicle
11. GPS Control Segment (OCX)*
12. Littoral Combat Ship
13. Global Combat Support System – Marine Corps
14. Small Diameter Bomb II
15. DDG 1000
16. CH-53K
17. Integrated Air and Missile Defense
18. Joint High-Speed Vessel*
19. Mission Planning System IV*
20. Multi-Platform Radar Technology Insertion Program*
21. EA-18G*
22. KC-46
23. E-2D AHE
24. P-8A
25. Net Centric Enterprise Services
26. Consolidated Afloat Networks and Enterprise Services
27. Integrated Strategic Planning and Analysis Inc 2
28. Cobra Judy Replacement*
29. NSA Key Management Infrastructure
30. NSA Public Key Infrastructure

*= Conjectured MS C Date

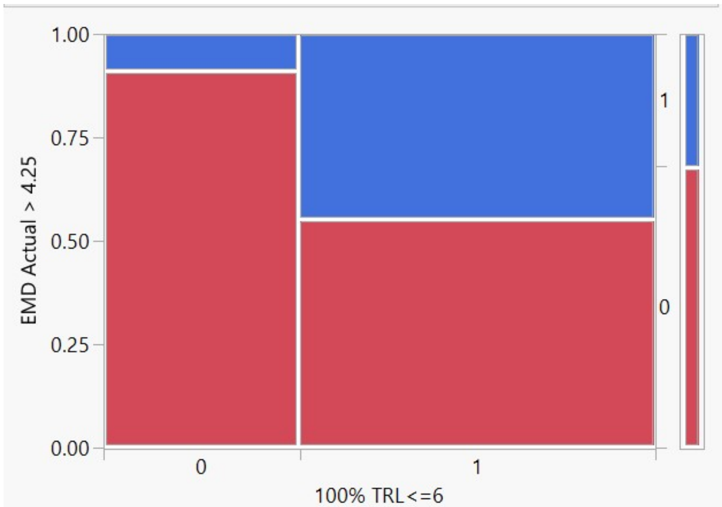
Appendix C – Bivariate Analyses

	RDTE	LN(RDTE)	Proc	LN(Proc)	Total	LN(Total)	EMD Actual	EMD Slippage
Cum Funding RDTE MS B	0.259	0.2097	0.8835	0.7846	0.8276	0.7271	0.1501	0.9072
% RDTE Funding at MS B	0.696	0.7174	0.047	0.064	0.2856	0.2231	0.1449	0.9373
% RDTE of Total Estimate at MS B	0.7955	0.7115	0.249	0.3227	0.8727	0.8082	0.1914	0.457
% Years Funded at MS B	0.8586	0.8018	0.0206	0.0464	0.3454	0.3378	0.3511	0.8543
MS B Start Year	0.8728	0.8064	0.1905	0.2041	0.7803	0.8974	0.7563	0.6754
MS B >2009	0.7028	0.7997	0.1883	0.2092	0.6474	0.7443	0.4098	0.6673
Est EMD Length	0.017	0.0137	0.1966	0.2442	0.0965	0.0947	0.0001	0.7092
EMD Length Actual	0.0001	0.0001	0.3937	0.5849	0.0644	0.0713		
EMD Slippage	0.0037	0.0062	0.43	0.2418	0.6215	0.6918		
Proc Qty MS B	0.4908	0.4733	0.4165	0.5348	0.2565	0.2267	0.7797	0.5465
Proc Est MS B	0.2144	0.1741	0.725	0.6352	0.8214	0.7333	0.0491	0.8935
RDTE Est MS B	0.1529	0.1178	0.6449	0.5795	0.6515	0.554	0.0213	0.7683
Total Estimate MS B	0.2068	0.1676	0.7162	0.6287	0.7996	0.7093	0.0423	0.8743
EMD % of Program Length	0.1104	0.1604	0.321	0.1722	0.9591	0.7993		
DV_Army	0.3323	0.4014	0.3784	0.3883	0.4589	0.4067	0.3023	0.2891
DV_Navy	0.6969	0.8322	0.2059	0.2772	0.7934	0.9175	0.9734	0.8315
DV_DoD	0.8325	0.7593	0.9065	0.7001	0.9451	0.8945	0.5887	0.8899
DV_AF	0.6296	0.5292	0.0424	0.0423	0.4434	0.4148	0.7145	0.6275
DV_Aircraft	0.8173	0.6798	0.8071	0.7733	0.6234	0.5632	0.1824	0.5456
DV_Ship	0.3652	0.3789	0.2599	0.3295	0.0909	0.045	0.251	0.95
DV_MAIS	0.0622	0.0384	0.5604	0.7986	0.2331	0.3064	0.0091	0.1754
DV_Vehicle	0.4906	0.4854	0.894	0.9848	0.6365	0.6636	0.6756	0.4678
DV_Satellite	0.0887	0.1001	0.9864	0.8566	0.0404	0.0768	0.9815	0.8439
DV_Missile	0.0002	0.0014	0.8787	0.9981	0.3568	0.3173	0.0041	0.0001
DV_Munition	0.6324	0.63	0.3443	0.4011	0.3405	0.2973	0.8092	0.0038
DV_Helicopter	0.1384	0.1101	0.0558	0.1374	0.0725	0.0894	0.0015	0.1219
DV_Boeing	0.7491	0.7754	0.2245	0.2416	0.6255	0.5705	0.7701	0.0646
DV_NorthropGrumman	0.0626	0.112	0.9322	0.692	0.0744	0.0981	0.3062	0.4237
DV_LockheedMartin	0.2783	0.3449	0.9046	0.8289	0.9874	0.8199	0.1531	0.6282
DV_GeneralElectric	0.2556	0.1826	0.0034	0.0293	0.0162	0.0261	0.0514	0.6366
DV_Raytheon	0.2479	0.3299	0.4045	0.5884	0.7521	0.8145	0.2196	0.0006
DV_JointContract	0.0847	0.1046	0.0831	0.1168	0.1327	0.1266	0.0038	0.3583
Concurrency	0.8305	0.9099	0.434	0.3922	0.5961	0.1143	0.4826	0.6779
DV_ProgComplete	0.1717	0.1128	0.6269	0.7477	0.8049	0.7432		
MS A?	0.8286	0.9089	0.9558	0.607	0.9548	0.9607	0.1395	0.9302
MS C Complete?	0.0113	0.0101	0.4973	0.3336	0.7344	0.557	0.0192	0.2868

% Program Complete by Funding	0.9595	0.8685	0.0153	0.0122	0.208	0.1143	0.9801	0.6162

	RDTE	LN(RDTE)	Proc	LN(Proc)	Total	LN(Total)	EMD Actual	EMD Slippage
Low TRL	0.835	0.7495	0.8435	0.9635	0.9842	0.9218	0.1579	0.9686
High TRL	0.9766	0.8798	0.2447	0.3442	0.291	0.3816	0.0606	0.3155
Mean TRL	0.7079	0.746	0.3365	0.4273	0.6283	0.667	0.0527	0.4729
% TRL Below 6	0.9891	0.8305	0.5435	0.4393	0.6984	0.5345	0.2642	0.4199
% TRL 6	0.4993	0.6268	0.2407	0.2353	0.5844	0.5149	0.0926	0.1634
% TRL 7	0.4224	0.423	0.9236	0.985	0.7899	0.7939	0.0772	0.5845
% TRL 8	0.8849	0.9722	0.29	0.346	0.5548	0.5332	0.525	0.4227
% TRL 9	0.9625	0.8757	0.2822	0.3762	0.4735	0.4849	0.2138	0.61
% TRL <=6	0.5142	0.5838	0.3552	0.3792	0.6914	0.6766	0.04	0.2867
% TRL > 6	0.5142	0.5838	0.3552	0.3792	0.6914	0.6766	0.04	0.2867
Low TRL <7	0.8106	0.9284	0.1232	0.165	0.3595	0.3212	0.2708	0.2106
Hight TRL <7	0.9367	0.9666	0.9024	0.9137	0.5883	0.7894	0.0427	0.3528
Mean TRL <6.3	0.3568	0.4114	0.4622	0.4693	0.8598	0.7378	0.0184	0.34
>73.9% TRL<=6	0.2753	0.3157	0.8995	0.9528	0.5979	0.5221	0.0354	0.4176
100% TRL <=6	0.9367	0.9666	0.9024	0.9137	0.5883	0.7894	0.0427	0.3528

Appendix D – Contingency Tables



Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9958	Prob(EMD Actual > 4.25=1) is greater for 100% TRL<=6=0 than 1
Right	0.0458*	Prob(EMD Actual > 4.25=1) is greater for 100% TRL<=6=1 than 0
2-Tail	0.0550	Prob(EMD Actual > 4.25=1) is different across 100% TRL<=6

Figure 9: 100% TRL <=6 against EMD Actual >4.25

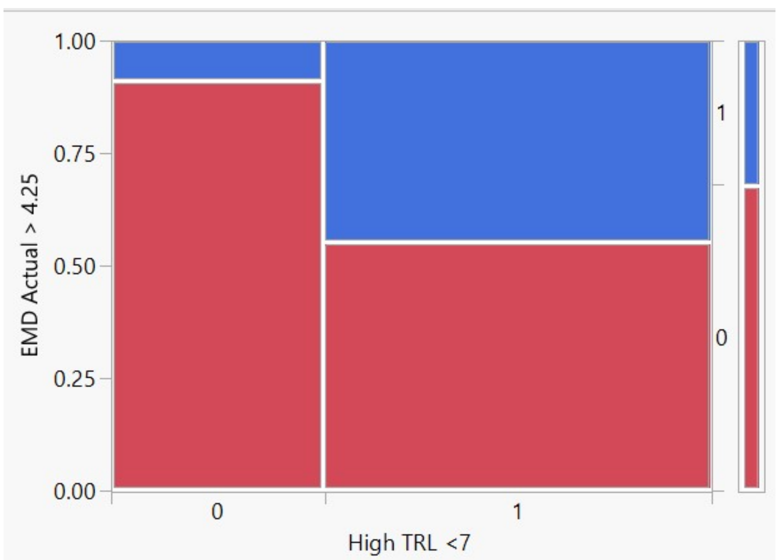


Figure 10: High TRL < 7 against EMD Actual > 4.25

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9958	Prob(EMD Actual > 4.25=1) is greater for High TRL <7=0 than 1
Right	0.0458*	Prob(EMD Actual > 4.25=1) is greater for High TRL <7=1 than 0
2-Tail	0.0550	Prob(EMD Actual > 4.25=1) is different across High TRL <7

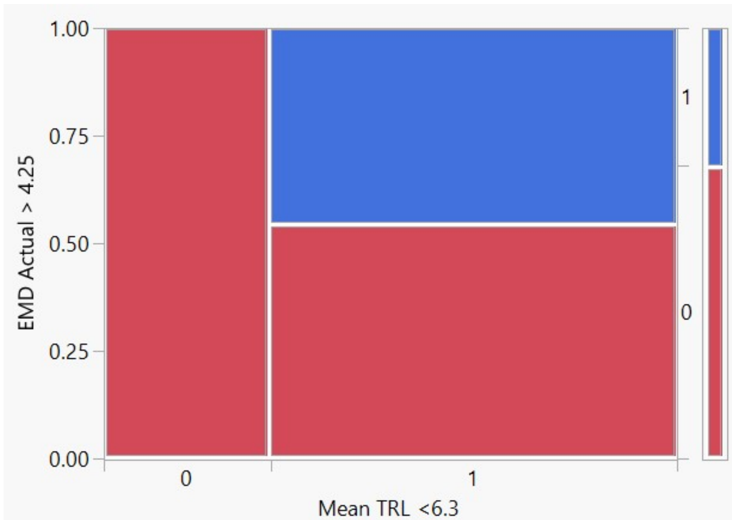


Figure 11: Mean TRL < 6.3 against EMD Actual > 4.25

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	1.0000	Prob(EMD Actual > 4.25=1) is greater for Mean TRL <6.3=0 than 1
Right	0.0146*	Prob(EMD Actual > 4.25=1) is greater for Mean TRL <6.3=1 than 0
2-Tail	0.0297*	Prob(EMD Actual > 4.25=1) is different across Mean TRL <6.3

Bibliography

Arena, Mark V., Robert S. Leonard, Sheila E. Murray, & Obaid Younossi, (2006). Historical Cost Growth of Completed Weapon System Programs. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/technical_reports/TR343.html

Birchler, Donald, Gary Christle, & Eric Groo, (2011). *Cost Implications of Design/Build Concurrency*. Ft. Belvoir, VA.: Defense Acquisition University.

Cancian, M. F. (2010). Cost growth: Perception and reality. *Defense Acquisition Review Journal*, 17(3), 389-404.

DeNeve, A. J., Ryan, E. T., Ritschel, J. D., & Schubert Kabban, C. (2015). Taming the hurricane of acquisition cost growth - or at least predicting it. *Defense Acquisition Research Journal,* 22(1), 84-105.

Department of Defense, Director, Research Directorate. (2009). *Technology Readiness Assessment Deskbook*.

Drezner, J. A., Jarvaise, J. M., Hess, R.W., Hough, P. G., & Norton, D. (1993). *An analysis of weapon system cost growth* (MR-291-AF). Santa Monica CA: RAND.

Dubos, G., Saleh, J., & Braun, R. (2008). Technology readiness level, schedule risk and slippage in spacecraft design: Data analysis and modeling. *Journal of Spacecraft and Rockets*, 45(4), 836–842. doi:10.2514/6.2007-6020

Dunbar, Brian. "Technology Readiness Levels Demystified." NASA, NASA, Aug. 2017, Retrieved from www.nasa.gov/topics/aeronautics/features/trl_demystified.html.

Dunbar, Brian. "Technology Readiness Level." NASA, NASA, Oct. 2012, Retrieved from www.nasa.gov/directorates/heo/scan/engineering/technology/txt_accordion1.html.

GAO. "Technology Readiness Assessment Guide." Gao.gov, Jan. 2020, Retrieved from www.gao.gov/assets/710/703694.pdf.

Hilmer, C. E., & Hilmer, M. J. (2014). *Practical econometrics: Data collection, analysis, and application*. New York, NY: McGraw Hill.

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

Jimenez, Christopher A. (2016). "Predicting Schedule Duration for Defense Acquisition Programs: Program Initiation to Initial Operational Capability". Masters thesis, Air Force Institute of Technology.

Katz, D.R., Sarkani, S., Mazzuchi, T., & Conrow, E.H. (2015). The relationship of technology and design maturity to DoD weapon system cost change and schedule change during engineering and manufacturing development. Systems Engineering, 18(1), 1-15. doi:10.1111/sys.21281

Kozlak, Scott J. (2016). "Predicting Cost Growth Using Programs Reviews and Milestones for DoD Aircraft". Masters thesis, Air Force Institute of Technology.

Light, Thomas, Robert S. Leonard, Julia Pollak, Meagan L. Smith, & Akilah Wallace, (2017). Quantifying Cost and Schedule Uncertainty for Major Defense Acquisition Programs (MDAPs). Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RR1723.html.

Linick, James. (2017). Technology Readiness Level (TRL) vs. Percent Development Cost. Presented at the 2017 ICEAA Professional Development & Training Workshop, 6-9 June 2017, Portland, OR.

McNicol, D. L. (2004). *Growth in the Costs of Major Weapon Procurement Programs*, Institute for Defense Analyses, IDA Paper P-3832.

Mankins, J. C. (1995, April 6). Technology Readiness Levels: A White Paper. Retrieved from https://aiaa.kavi.com/apps/group_public/download.php/2212/TRLs_MankinsPaper_1995.pdf.

Smoker, R.E., & Smith, S. (2007). System cost growth associated with Technology Readiness Level. *Journal of Parametrics*, 26(1), 8-38. doi:10.1080/10157891.2007.10462276.

Younossi, O., Arena, M., Leonard, R., Roll, C., Jain, A., & Sollinger, J. (2007). *Is Weapon System Cost Growth Increasing? A quantitative assessment of completed and ongoing programs*, RAND Corporation, doi:10.7249/mg588

Zach. (2020, December 31). How to perform a Breusch-Pagan test in excel. Retrieved March 1, 2021, from https://www.statology.org/breusch-pagan-test-excel

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
25-03-2021	Master's Thesis	August 2019 – March 2021

TITLE AND SUBTITLE

AN ANALYSIS OF EFFECTS OF TECHNOLOGY READINESS LEVELS ON COST GROWTH

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Bissing, Christopher R., First Lieutenant, 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/EN)
2950 Hobson Way, Building 640
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENV-MS-21-M-207

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AIR FORCE LIFECYCLE MANAGEMENT CENTER
1865 Fourth St, Bldg 14, WPAFB, OH 45433
937-656-5504 and shawn.valentine@us.af.mil
ATTN: Shawn Valentine

10. SPONSOR/MONITOR'S ACRONYM(S)

AFLCMC/FZC

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

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION 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

This research seeks to evaluate the effects of Technology Readiness Levels (TRL) on Cost Growth. It makes use of data from Technology Readiness Assessments (TRA) and Selected Acquisition Reports (SAR) to explore relationships between TRLs at Milestone B and cost growth in Major Defense acquisition Programs (MDAP) and Major Automated Information Systems (MAIS). Programs using higher proportions of critical technologies rated below TRL 7 tend to experience greater cost growth than programs that use more mature technologies. Current DoD doctrine requires TRL 6 to enter Milestone B. The results of this research seek to evaluate the merit of this requirement. TRL usefulness in multiple linear regression models is assessed by comparing against regression models without the use of TRLs. Results indicate relationships between TRL and Cost growth may be driven by omitted variables such as length of EMD phase. A more complete dataset may indicate that TRLs are driving EMD length and provide insight into potential causes of schedule slippage.

15. SUBJECT TERMS
Technology Readiness Levels, EMD Cost Growth, EMD Schedule

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Dr. Edward D. White AFIT/ENC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	76	(937) 255-3636 x 4540 Edward.white@afit.edu

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