3-2004

Estimating Cost Growth in Engineering and Schedule Cost Categories Using a Two-Pronged Regression Approach

Chris J. McDaniel

ESTIMATING COST GROWTH IN ENGINEERING AND SCHEDULE

COST CATERGORIES USING A TWO-PRONGED REGRESSION APPROACH

THESIS

Chris J. McDaniel, Captain, USAF

AFIT/GCA/ENC/04-03

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GCA/ENC/04-03

ESTIMATING COST GROWTH IN ENGINEERING AND SCHEDULE COST CATERGORIES USING A TWO-PRONGED REGRESSION APPROACH

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

Chris J. McDaniel, MBA

Captain, USAF

March 2004

AFIT/GCA/ENC/04-03

ESTIMATING COST GROWTH IN ENGINEERING AND SCHEDULE COST CATERGORIES USING A TWO-PRONGED REGRESSION APPROACH

Chris J. McDaniel, MBA
Captain, USAF

Approved:

//Signed//	8 Mar 04
Edward D. White (Chairman)	date
//Signed//	8 Mar 04
Michael A. Greiner, Major, USAF (Member)	date
//Signed//	8 Mar 04
Michael J. Seibel (Member)	date

AFIT/GCA/ENC/04-03

Abstract

Cost growth in major DoD acquisition programs has been commonplace for the last 35 years, and shows no signs of improvement despite the adoption of new business practices legislation. In the current environment where taxpayer dollars are heavily competed for, and the expenditure of those dollars is highly scrutinized, it has become a high priority in Department of Defense leadership to build accurate cost estimates that reduce overruns and restore credibility to the defense acquisition process.

Previous research has validated the use of two-pronged logistic and multiple regression approach that offers better predictive ability than the traditional multiple regression approach alone. This research further validates the use of this two-pronged approach by applying it to the engineering and schedule cost growth categories.

We update and augment previously collected programmatic data from the Selected Acquisition Reports (SARs) between 1990 and 2001 for programs covering all defense departments, with the latest SAR database (1990-2002). We start the analysis by building logistic regression models to predict whether cost growth will occur. Then we build multiple regression models to predict the extent to which a program will experience cost growth. The response variables for our models are the respective cost growth on procurement-funded efforts in the engineering and schedule cost categories, during the Engineering and Management Development phase of the acquisition process.

Acknowledgements

First, I would like to thank my thesis Advisor, Dr. Edward White III, for sparking my interest in regression and providing support when things got confusing (I kept him busy). Next, I would like to thank my academic advisor, Major Michael Greiner for his leadership and understanding during a challenging 18 months. I would also like to thank Mr. Michael Seibel for his interest and inputs to our thesis process.

Next, I thank my thesis partner and friend, 1Lt Matt Rossetti for keeping me focused when I started to procrastinate, but more importantly, for smoking a good cigar with me on special occasions (like Tuesdays).

I would especially like to thank my wife Tina, for her unwavering support and love during the last few months of late nights and working on weekends. She handled being an AFIT widow like a champ (most of the time). Finally, I give special thanks to my daughter Megan, for watching Spongebob Squarepants with me when I needed a good laugh.

Chris J. McDaniel

Table of Contents

Page

List of Figures

List of Tables

Table Page

ESTIMATING COST GROWTH IN ENGINEERING AND SCHEDULE COST CATERGORIES USING A TWO-PRONGED REGRESSION APPROACH

I. Introduction

General Issue

Defense spending has undergone great change in the last 20 years. During the Reagan Administration of the 1980s, the Cold War saw high levels of defense spending. In 1985, the United States spent over $245 billion for national defense, a significant 25.9% of the President's Budget (OMB, 2004: 73, 78). The arms race with the former Soviet Union kept funding for weapon system acquisition flowing with relative ease.

As time passed, however, defense spending became heavily scrutinized as public perception of waste and excessive funding grew. In the years following the Cold War, particularly under the Clinton Administration of the 1990s, the United States experienced record-setting reductions in defense spending. By 2002, the budget for national defense hovered around $332 billion, a mere 16.5% of the President's Budget (OMB, 2004: 75, 80).

Unfortunately, global threats to the security of the United States have not declined in the past 20 years, merely changed form. This puts the defense acquisition community in the position of having to find ways to do more with less. For this reason, elected representatives, as well as higher ranking members of the Department of Defense pay close attention to the cost performance of major defense acquisition programs (MDAPs). With each new administration, a movement to reform the Department of Defense's

(DoD) major acquisitions programs and processes begins. This movement has gained serious momentum over the past decade. Major weapon systems being completed over budget and behind schedule is the motivation behind the current movement.

Cost growth in the procurement of major weapon systems can be attributed to poor program management or contractor inefficiencies, however, it mainly stems from risk and uncertainties about the program (Bielecki, 2003:2). In a 1993 RAND study, Drezner and others sought to characterize cost growth (variance between initial and final contract baselines) against a wide variety of factors. In general, they found that during the time period between McNamara's reforms (1965) and 1990, cost growth hovered at around 20 percent, on average. In the last 15 years, the DoD has seen more reforms such as the Packard Commission of 1986, the Goldwater-Nichols Act of 1987, and the Acquisition Reform movement. In spite of claims that these reforms would lead to cost reductions, Air Force cost overruns grew another 9.9 percent (Suddarth, 2002:7). This 29.9 average cost growth is confirmed by the Assistant Secretary of the Air Force (Acquisition), Dr. Marvin Sambur, and the Deputy Chief of Staff for Air and Space Operations, Lieutenant General Ronald Keys, during their statement before the House Armed Services Committee on April 2, 2003 where they stated that for the Air Force, program execution problems had resulted in average cost growth of 30% for acquisition programs (Sambur/Keys, 2003).

In order for the DoD to retain its' credibility with Congress and the American taxpayer, this cost growth must be slowed, contained, and reduced. DoD program managers must concern themselves with accurately identifying the cost risks associated with potential cost increases in their program cost estimates. To control cost growth, managers must focus on accurately assigning dollar values to risks, so that the original estimate from which we calculate cost growth is more accurate (Bielecki, 2003:2)

Specific Issue

The primary objective of weapon system cost estimating provides decision makers with an accurate estimate of the resources required to complete a project. To this end, cost estimators have many methodologies at their disposal: analogy, engineering, actual, and parametric.

The highly subjective analogy method compares a new system with an existing system for which there are accurate cost and technical data, and is most often used early in the program when little is known about the specific system being developed. Later in the program, the engineering estimate, commonly referred to as the "bottom up" method, is used when the scope of work is well defined and a comprehensive Work Breakdown Structure (WBS) is in place. Actual costs are used whenever they are available, but they are rarely available in the early stages of a program.

The parametric (statistical) method is used to analyze our data during this research. This method allows the cost estimator to objectively analyze large databases of historical data and make inferences about the relationship of the cost risk associated with one or more program parameters. The parametric technique is used early in the program to estimate cost risks throughout the life cycle of a program using statistical regression techniques to develop cost estimating relationships (CER).

Using regression to predict whether or not a program experiences cost growth, and the magnitude of that growth (should it occur) are the key focuses of this research. This study builds upon the thesis work of Bielecki (2003), Moore (2003), and Sipple

(2002) to provide the cost estimating community a model to accurately estimate cost risk of the *engineering* and *schedule* cost variance categories of the procurement appropriation during the EMD phase of defense acquisition programs.

Scope and Limitations of the Study

Fundamental to any discussion of cost growth is the Selected Acquisition Report (SAR); “Since 1969, Congress has required DoD to submit SARs on its major acquisition programs” (Calcutt, 1993:3). They are readily available and contain relatively reliable data on cost growth. As SARs are the foundation from which cost growth is analyzed, so too, they are our source of data for this study. The SAR contains the following three cost estimates useful for analyzing program cost growth:

- Planning Estimate (PE): This is the DoD estimate normally made during the Concept Exploration and Definition phase of the acquisition cycle (Calcutt, 1993:3).
- Development Estimate (DE): This is the estimate established at Milestone II, which begins the Engineering and Manufacturing (EMD) phase of the acquisition cycle (Calcutt, 1993:3).
- Current Estimate (CE): This is the most up-to-date estimate of what the program will cost at completion (Calcutt, 1993:3).

The SAR reports cost variances in base year and then year dollars (allowing for analysis between programs on a constant dollar basis) and classified into one of the following seven categories:

1. Economic: changes in price levels due to the state of the national economy
2. Quantity: changes in the number of units produced
3. Estimating: changes due to refinement of estimates
4. Engineering: changes due to physical alteration
5. Schedule: changes due to program slip/acceleration
6. Support: changes associated with support equipment
7. Other: changes due to unforeseen events

(Drezner, 1993:7)

The security classification of some of the programs will limit our research. Any program with a confidential or higher classification will not be looked at in this study. Given that this type of information is not classified as confidential or higher on the vast majority of Major Defense Acquisition Programs (MDAPs), this limitation is viewed as having negligible impact on the utility of the model we build. Other limitations exist within the SAR which are discussed further in Chapter 3.

For the purposes of this research, cost growth is measured as a positive percentage increase from the DE to the latest CE as reported in the SAR. Since we build upon the research previously fielded by Sipple, Bielecki, and Moore, we employ the same framework and methodologies initiated by Sipple and expanded by Bielecki and Moore. The difference being that this study focuses on the *engineering* and *schedule* cost variance categories of the procurement appropriation during the EMD phase of defense acquisition programs. In particular, this research builds logistic and multiple regression models with predictor variables from the EMD phase that predict whether or not a program experiences cost growth (logistic) and, if it exists, how much it experiences (multiple). Additionally, we utilize the database conceived by Sipple (2002), update it to contain the latest CE (2002 data) of each program, if applicable, and add any new programs that are at least three years into the EMD phase (mature program).

Research Objectives

The purpose of this research is twofold. First, logistic regression (yes or no response) will be used to ascertain if there are certain parameters within the program that are able to predict if a program will experience cost growth in the *engineering* and *schedule* cost variance categories of the procurement appropriation during the EMD phase of program development. Second, if cost growth is present, multiple regression will be used to determine how much growth occurs.

Chapter Summary

This research expands the cost estimating methodology originally developed by Sipple, and further developed by Bielecki and Moore. Our specific goal provides the cost estimating community an effective model to estimate the cost risk associated with a program early in its development, and the overall goal reduces the DoD cost growth rate from its current levels. We continue with Sipple's two step methodology — analyzing SAR historical data with logistical and multiple regression to successfully predict cost growth in the EMD of program development. In the following chapter we present an overview of the acquisition process and its' environment, examine cost risk and the effect it has on our study, and finally, investigate past research in cost growth.

II. Literature Review

Chapter Overview

This chapter establishes a historical framework from which to base our methodology and develop our models. First, we discuss the acquisition process, past and present, and how that process affects our approach in this study. Next, we look at the acquisition environment to familiarize ourselves with the increasing importance of these types of models. Cost risk and its considerations are addressed after the environment has been established. We conclude the chapter with a review of recent studies that have relevance to ours.

The Acquisition Process

Being that this research focuses on a very specific portion of the overall acquisition process, we begin this chapter with a brief overview of how that process works and where our focus lies. To this end, we start with Department of Defense Instruction (DoDI) 5000.2 Operation of the Defense Acquisition System, which "Establishes a simplified and flexible management framework for translating mission needs and technology opportunities, based on approved mission needs and requirements, into stable, affordable, and well-managed acquisition programs that include weapon systems and automated information systems." (DoDI 5000.2, 2003:1).

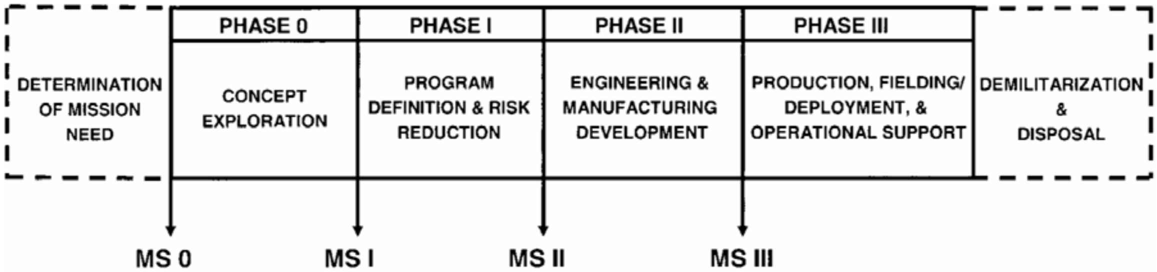


Figure 2.1 - Acquisition Milestones and Phases (DoDI 5000.2, 2000:1)

Figure 2.1 is a graphical representation of what the Defense Acquisition Management Framework looked like prior to a January 2001 change to DoDI 5000.2. We include this past business practice because the SAR data in our database is based on this format. The process consists of four milestones (MS 0-MS III) and four phases (PHASE 0-PHASE III), described below. This information was extracted from the DoD 5000.2, prior to the Jan 2001 change.

- *Approval to conduct concept studies (MS 0)*- The Milestone Decision Authority (MDA) approves short-term concept studies and the PHASE 0 exit criteria.
- *Concept Exploration (PHASE 0)*- Evaluate the feasibility of alternative concepts, determine the most promising concepts and solutions.
- *Approval to begin new acquisition program (MS I)*- MDA approves the Acquisition Strategy, Cost as an Independent Variable (CAIV) objectives, initial Program Management Baseline (APB) and PHASE I exit criteria.
- *Program Definition and Risk Reduction (PHASE I)*- Design the system, demonstrate critical processes and technologies, and develop prototypes.

- *Approval to enter Engineering and Manufacturing Development (EMD) (MS II)*- Approval of Acquisition Strategy, CAIV objectives, updated APB, Low-Rate Initial Production (LRIP) quantities, live-fire and Test and Evaluation (T&E) waiver (if applicable) and PHASE II exit criteria.
- *Engineering and Manufacturing Development (PHASE II)*- Mature and finalize selected design, validate manufacturing and production processes and test and evaluate the system.
- *Production or fielding development approval (MS III)*- Approval of Acquisition Strategy, production (weapon systems), deployment (information systems), updated APB and PHASE III exit criteria.
- *Production, Fielding or Deployment and Operational Support (PHASE III)*- Produce system, field it, monitor mission performance, support fielded system, modify or upgrade as required.

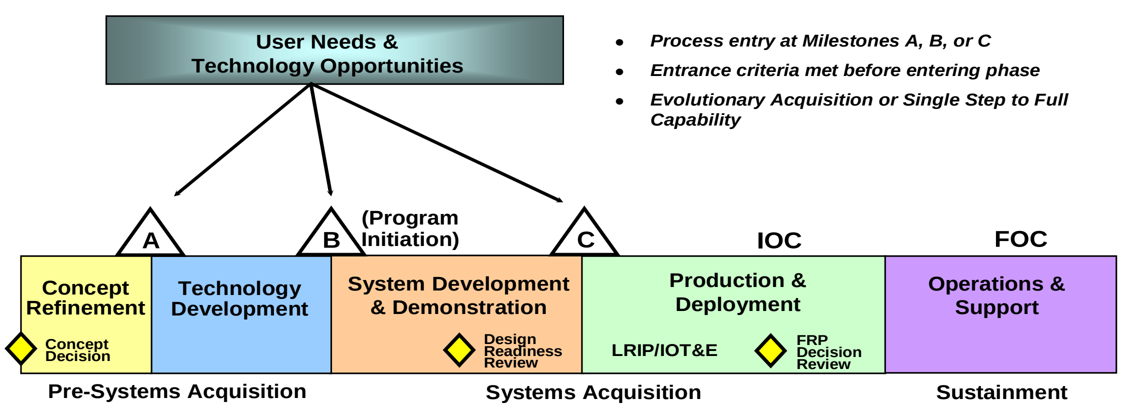


Figure 2.2 - Acquisition Milestones and Phases (DoDI 5000.2, 2001:1)

Figure 2.2 is a graphical representation of what the Defense Acquisition Management Framework looks like now, due to the aforementioned change to the DoDI

5000.2 in January of 2001. It replaces the traditional milestones with an ABC format and labels the phases by name (as opposed to numbering or lettering them). The following is a brief overview of the new framework, taken from the current DoD 5000.2.

- *Concept Refinement Phase*- Refine the initial concept and develop a Technology Development Strategy (TDS). This phase cannot begin until the MDA makes a Concept Decision and does not mean that a new acquisition program has been initiated.
- *Milestone A*- MDA approves the TDS.
- *Technology Development Phase*- Reduce technology risk and determine the appropriate set of technologies that will be integrated into the full system. This process is iterative in that it assesses the viability of available technologies and refines user requirements simultaneously.
- *Milestone B*- The acquisition program has officially started. For programs using Evolutionary Acquisition (which will be described in more detail later in this chapter), each increment will have its own Milestone B. This is where the PM and MDA prepare and approve an Acquisition Strategy.
- *System Development and Demonstration*- Develop full or increment of capability, reduce integration and manufacturing risk, ensure operational supportability, implement human systems integration, and design for producibility.
- *Milestone C*- MDA commits the DoD to production and authorizes entry into LRIP, production and limited deployment for operational testing.

- *Production and Deployment Phase*- Achieve operational capability that satisfies mission needs, either incrementally or fully.
- *Operations and Support Phase*- The two major components of this phase are sustainment and disposal. The purpose being to ensure the system continues to perform its mission and is ultimately disposed of properly.

As you can see, we did not go into as much detail on the new acquisition framework as we did on the old. The reason for this is simple: our study is based on the old phases and milestones because all of our historical data (from the SARs) is based on the old process. It is also important to note at what point we focus on in the acquisition process.

Acquisition Timeline:

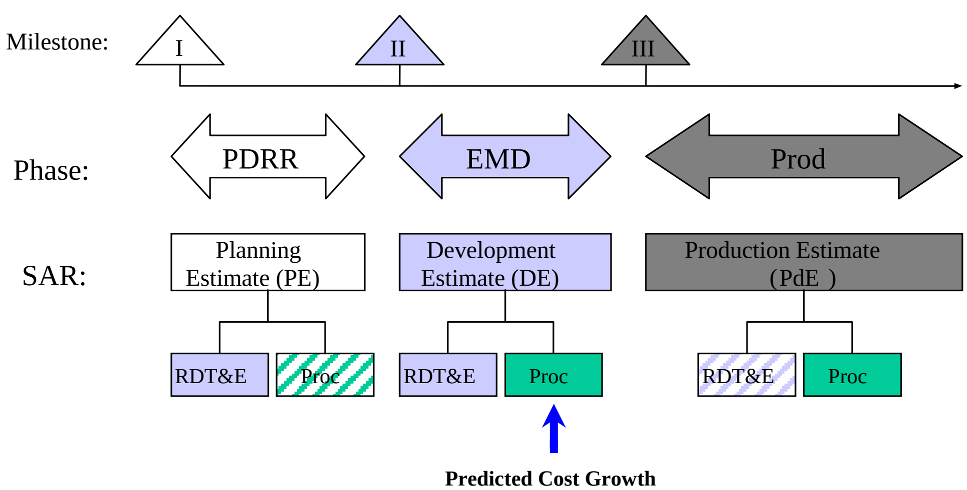


Figure 2.3 - Acquisition Timeline (Dameron, 2001:4)

Later in this chapter, we review the thesis work on this subject of our predecessors (Sipple, Bielecki and Moore). Sipple focuses on the *engineering* cost variance (CV) category and Bielecki on the *estimating, schedule, support,* and *other* categories of the RDT&E appropriation. While these studies target specific CV categories, Moore targets the overall procurement appropriation in the EMD phase. Our research focuses on the individual CV categories of *engineering* and *schedule*. We make the assumption that the cost estimator is more concerned with specific areas of cost growth.

The Acquisition Environment

The acquisition process is under great scrutiny as evidenced by the sweeping changes in the overall acquisition framework in January of 2001. The changes, however, do not stop there. The latest initiative to revamp the current acquisition process is traced back to September 2002 when the Secretary of Defense issued an unsigned memorandum stating that the current regulations were "overly prescriptive and do not constitute an acquisition policy environment that fosters efficiency, creativity and innovation." As a result, said the memo, the 5000 series, which includes versions 5000.1 and 5000.2, would be "cancelled ... effective immediately." (Erwin, 2002)

On 12 May of this year (2003), DoD Directives 5000.1 and 5000.2, were signed by the Deputy Secretary of Defense and replaced the same directives previously dated October 23, 2000. One of the policies instituted by this directive is that of cost and affordability:

> All participants in the acquisition system shall recognize the reality of fiscal constraints. They shall view cost as an independent variable(CAIV), and the DoD Components shall plan programs based on realistic projections of the dollars and manpower likely to be available in future years (DoD Directive 5000.1, 2003:4).

This policy indicates the importance of CAIV to program management and signifies the extent to which the OSD believes cost estimation should be used in budgeting. Realistic projections become extremely important in that appropriated funds are scarce and under heavy supervision by multiple stakeholders. In addition, when taken into account the number of government civil servants, military officers and enlisted troops that it takes to make funding changes, it is fair to assume that administrative costs due to poor planning are high, and could be reduced with more accurate initial estimates. For these reasons, each program manager must strive to get their cost estimations right, more often than not, so they can maintain their programs' credibility with DoD executives, Congress, and the American public.

The seriousness of this acquisition reform movement is echoed yet again in April 2003 when Dr. Marvin Sambur, Assistant Secretary of the Air Force (Acquisition), and the Deputy Chief of Staff for Air and Space Operations, Lieutenant General Ronald Keys, state before the House Armed Services Committee:

> In the past, we have designed our programs with a 60-70% confidence level of meeting cost, schedule, and performance goals. In order to be credible both to the warfighters and Congress, I have implemented a 90% confidence level in meeting our requirements. By demanding collaboration between all the parties, we can ensure the right trade-offs are made throughout the acquisition process to meet the required goals. It is imperative that, both the warfighting and acquisition communities work together to make tradeoffs of non-critical elements within programs to buy down risk, throughout the acquisition cycle. Bottom line: credibility means delivering what we promise, on time and on budget (Sambur/Keys, 2003).

Clearly, a major concern in the acquisition community is that of credibility and fiscal responsibility. And it would be difficult to have one and not the other. To obtain this credibility, the pressure is on the cost estimator to accurately predict the costs associated with the program at all phases of the system life cycle. This is no easy challenge. The methods available to the estimator range from subjective methods (quick and easy) to objective methods (time consuming and complex), both of which have their strengths and weaknesses, and both must address risk.

Cost Risk

"Risk: Minimizing the possibility that something goes wrong" (Cancian, 1995:191). Cancian's definition may appear oversimplified, but it's a great place to start. As cost estimators, much of the risk we encounter involves uncertainty. Uncertainty about the countless variables we have identified, and uncertainty about the variables we fail to identify. These uncertainties have great potential to make "something go wrong" in our estimates. This is especially true when attempting to estimate the cost of a system that has not yet been built.

A cost estimator must first identify and consider all areas of uncertainty associated with that system and related future events. Once identified and estimated, the cost risk is translated into a dollar figure which can then be used by decision makers. The Air Force Materiel Command (AFMC) *Financial Management Handbook* confirms "program risk refers to the uncertainties and consequences of future events that may affect a program", and goes on to say that "risk is the summation of probable effects of unknown elements in technical, schedule, or cost related activities within the program." The latter of these three risk parameters asks the question: "can the program as presently structured technically and with respect to schedule, be completed for the budgeted amount of money?" (*AFMC Financial Management Handbook,* 1998:11-20)

In the case of the Air Force's most expensive acquisition program, the Advanced Tactical Fighter (a.k.a. the F-22 Raptor), the answer to this question has historically been "no". This program is an excellent example of how uncertainty creates risk. Although there are countless factors (especially in the EMD phase) that can be held responsible for F-22 program cost growth, a very interesting uncertainty is worth mentioning. According to a 1999 GAO report, "A factor the Air Force did not consider in its estimate of potential cost growth was the possibility that the F-22 program may have to absorb a higher share of the manufacturing plant's overhead costs if the contractor does not sell enough C-130J aircraft, which are produced at the same plant as the F-22." (GAO/NSIAD-99-55, 1999:5). Ironically, this is a factor that the Air Force would have easily been able to predict (since C-130J is also a DoD acquisition program) had they realized its potential impact on cost growth.

The F-22 program is also an excellent example of what could be argued is a program's biggest risk of all: being cut. Funding instability is a fact of life that the F-22 has been dealing with for years. This is because "as threats began to change, developmental challenges arose, and total ownership costs continued to mount, it was unlikely to be overlooked as a prime source of funding for other 'must pay' bills." (Myers, 2002:322). The truth of this statement is easily reflected in the Defense Subcomittee's rationale behind their $1.8B cut in the 2000 Department of Defense Appropriations Bill:

> It is clear from a larger perspective, the F-22 is consuming resources that could be used to address other critical strategic concerns such as emerging threats from chemical/biological/nuclear terrorism, information warfare, and cruise missiles. (Defense Subcommittee, 2000)

The bottom line is that a cost analyst must deal with countless unforeseen events in order to protect their program's funding, and thus, the program itself. The AFMC Financial Management Handbook discusses three methods the analyst can use to approximate the likelihood of a certain event occurring: *a posteriori*, (after the fact), *a priori* (a prediction based upon theoretical probability distributions), or *subjective judgment* (*AFMC Financial Management Handbook*, 1998:11-21). No matter which method the estimator chooses, the end product will depend largely on the skill of the estimator, the level of accuracy required, the level of detail needed, and the time required (and available) to complete the estimate. These are also the factors that will determine how well an analyst mitigates risk when applying their chosen methodology.

We mentioned in Chapter 1 that the cost estimating community has different cost estimating methodologies at their disposal including, but not limited to, analogy, engineering, actual, and parametric. These methods are widely accepted and practiced in both the DoD and civilian sectors. Figure 2.4 shows the techniques recognized by the Ballistic Missile Defense Organization (BMDO) cost estimating community. These techniques are also widely accepted and practice in most cost estimating communities. It is interesting to note that as the level of detail and difficulty of gathering the data increase, the level of precision tapers off.

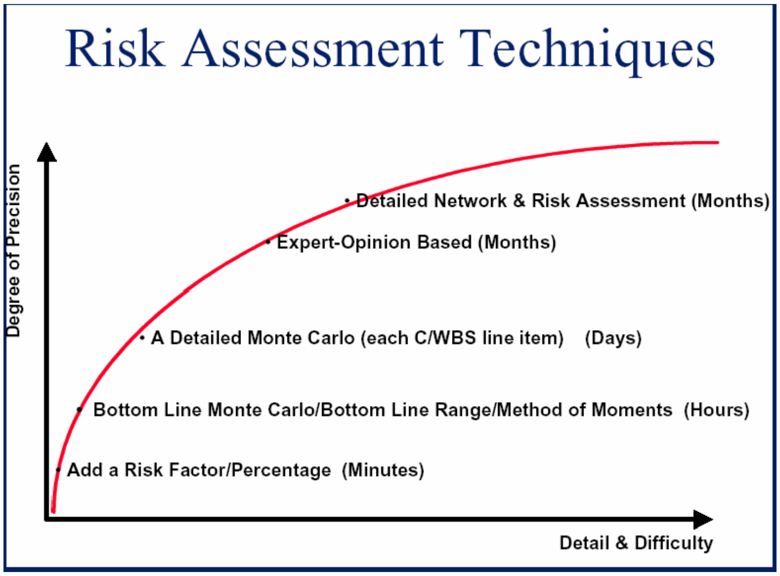


Figure 2.4 - Risk Assessment Techniques (Coleman, 2000:4-9)

In conclusion, risk needs to be addressed up front and early and the cost estimator's role in this process is crucial. This philosophy is made very clear by the Air Force Materiel Command (AFMC) *Financial Management Handbook*:

> Because resources are limited, considerable time and effort in planning for future acquisitions is necessary. The central issue in such planning usually concerns resource allocation. Cost analysis supports acquisition decisions required to allocate financial resources among alternative systems. The acquisition process revolves around the cost estimate - budgets are based on estimates and future cost performance is measured against estimates. Cost estimating must be accurate if the operation of the Planning, Programming, and Budgeting System (PPBS) is to be realistic, and effective decision making is to take place (*AFMC Financial Management Handbook*, 1998:11-2)

Past Research in Cost Growth

A benefit to doing continuing research on three comprehensive studies on cost growth is that the previous authors: Sipple, Bielecki, and Moore, provide us with an exhaustive review of the pertinent literature on cost growth from 1974 through 2001.

Sipple's review of the literature was thorough enough that the follow-on work performed by Bielecki and Moore provides us with no relevant studies outside of their own findings. The important thing to note here is that the unique two-step methodology borne by Sipple to identify and then quantify cost growth is tangent to existing studies on predicting cost growth.

Sipple provides us with twelve relevant studies on this matter, see Table 2.1. For a complete review of the studies listed refer to Sipple (2002). These studies influenced Sipple in his development and creation of the predictor variables used in both the logistical and ordinary least squares (OLS) models.

Author (Year)
IDA (1974)
Woodward (1983)
Obringer (1988)
Singleton (1991)
Wilson (1992)
RAND (1993)
Terry & Vanderburgh (1993)
BMDO (2000)
Christensen & Templin (2000)
Eskew (2000)
NAVAIR (2001)
RAND (2001)

Table 2.1 - Sipple Thesis (Sipple, 2002:20-44)

Sipple Thesis

Where Sipple's methodology differs from previous studies is that Sipple looks at predicting cost growth in the EMD phase of the system life cycle instead of attempting to predict overall cost growth for an entire system life cycle. This approach affords us the ability to break down the cycle into its different phases: PDRR, EMD, and Prod and further into the appropriations contained in each and study the effects that over 75 predictor variables have on these appropriations given a particular phase. Sipple is also unique in that he recognizes that the Y response variable (*Engineering percent*) exhibits a mixed distribution. "About half of the distribution is continuous, while the other half is massed at one value, zero—indicating no cost growth. This mixed distribution scenario generally calls for splitting the data into two sets" (Sipple, 2002:58). We will utilize these same variables and two-step methodology in our approach to predict cost growth in the production phase given data from the EMD phase.

The goal of Sipple's research is to predict cost growth in the EMD Phase as it relates to RDT&E appropriations in the SAR engineering cost variance category. Sipple collects SAR data and builds a database of over 75 predictor variables using 115 major acquisition programs. He then uses logistic regression to first identify if cost growth exists. If it exists, OLS regression is implemented to indicate how much cost growth will occur. "Sipple demonstrates through the use of four regression models (A, B, C, D) that the combination of logistic and multiple regression produce similar predictive results as a traditional single-step multiple regression cost estimating methodology. However, the two-step methodology is preferred to the single-step methodology because of the stronger statistical foundation achieved with the two-step method" (Bielecki, 2003:21).

The first of the four models, Model A, uses logistic regression to predict whether or not a program will have cost growth (yes or no). Programs with positive cost growth are given a response of "1" in the database, while programs with negative cost growth are given a response of "0". Model A uses the predictor variables to regress the "0/1" response on 80 percent (90 data points) of the data points (the remaining 20 percent are used to validate the models). Model B uses OLS to predict the amount of cost growth that will occur. Sipple builds Model B to regress the same variables on 47 of the 90 data points that were found to contain positive cost growth in Model A. This time the response is the percentage of actual cost growth (instead of 0/1) and utilizes a log transformation on the Y response to correct for non-constant variance (heteroscedasticity) of the residuals. Model C also uses OLS to predict the amount of cost growth, but does not correct for this heteroscedasticity and subsequently fails the OLS statistical assumptions of normality and constant variance (homoscedasticity) rendering it ineffective for drawing statistical conclusions about the amount of cost growth. These A and B models represent the two-step method.

Sipple builds Model D to test his two-step method against a single step or single model method. Model D does not employ logistic regression. Instead, OLS is used on the entire 90 data points and the response is not transformed. Again, the underlying assumptions of normality and constant variance are not met without the log transformation. Instead of predicting a “0” or “1” response, Model D will predict both negative and positive values. In this case, programs with no cost growth will have a predictive response of zero or less.

Upon validation of the four models using the 20 percent test set, Sipple found that both Models A and B accurately predicted the existence of cost growth and the amount of cost growth with about a 70 percent accuracy rate. Model A utilizes seven out of 78 possible predictor variables, while Model B uses three. Model C does fairly well at predicting the validation data. Using an 80 percent confidence bound, Model C contains

73 percent of the data, however, due to the violation of the OLS assumptions, it is unknown whether or not this confidence bound is a true 80 percent.

Model D's predictive ability is comparable to Model A. Model A predicts correctly 66.06 percent of the validation points, while Model D predicts correctly 62.87 percent of the validation points. Comparing Model D to Model B, Sipple found that "Model B produces higher R^2 values than Model D…Model B yields more predictive ability for the number of variables, and none of Model D's versions can compare to the versions of Model B above two predictor variables" (Sipple, 2002:104).

It would appear that the two-step methodology employing Models A and B is superior than using a one model approach. The C and D Models seem to perform well, but their lack of conformity with underlying regression assumptions greatly reduces the ability of the user to accurately interpret their results (Sipple, 2002:113).

Bielecki Thesis

Employing the same methodology and underlying philosophy, Bielecki carries Sipple's work forward to research cost growth in the four remaining SAR cost variance categories: schedule, estimating, support, and other. Bielecki employs logistic and multiple regression to build models aimed at identifying cost growth characteristics in each category as they relate to RDT&E appropriations in the EMD phase of the system life cycle.

Bielecki also finds that the distribution for each cost growth category are mixed — indicating the need for the two-step approach. In addition, he observes that the *other* and s*upport* categories do not contain enough data to support a meaningful statistical analysis. Therefore, Bielecki limits his study to the remaining two categories: schedule and estimating.

As Sipple does before him, Bielecki builds a family of logistic and multiple regression models for each category and picks the best one for each. The best logistic regression model submitted for each category validates at 85.71 percent and 78.26 percent for the schedule and estimating categories respectively. Using an 80 percent confidence bound, the best multiple regression model submitted for each category validates at 80.00 percent and 100 percent for the *schedule* and *estimating* categories, respectively.

Moore Thesis

Unlike Sipple and Bielecki, Moore's research does not focus on a specific SAR cost variance category. Instead, Moore focuses on the procurement appropriation and any cost growth associated with it in the EMD phase of the system life cycle as he states this is the "next logical level" (Moore, 2003:16).

When Moore performs a preliminary analysis of his data, he found that the distribution for procurement cost growth during the EMD phase exhibits identical characteristics to those found by Sipple (Moore, 2003:21). Meaning that there is a mixed distribution and the two-step methodology will be used.

The logistic regression model Moore submits for validation accurately predicts four out of the four data points available for validation. Of the 25 data points randomly selected for validation, only four of them contained the variable *FUE-based Maturity*. Upon further validation, the model was found to accurately predict 37 out of the 39 data points used to build the model. Therefore, this logistic model is found to be highly predictive. The multiple regression model Moore submits for validation also accurately predicts 100 percent of the predicted data points, using an 80 percent prediction interval (Moore, 2003:47).

OSD CAIG Study

In addition to the above three theses, a study by the Office of the Secretary Defense Cost Analysis Improvement Group (OSD CAIG) is found to be relevant to our study and is therefore included in our literature review.

The study, *Cost Growth of Major Defense Programs,* is the culmination of 10 years of research between the OSD CAIG, NAVSHIPSO and AT&T. This study uses the SARs of 286 programs as its source of data. When bumped up against the study criteria: unclassified, milestone II captured, three years of data past milestone II, and data complete; these 286 programs are reduced to 142 and are entered into the database.

They define cost growth as the “difference between today’s estimate and a baseline estimate caused by:”

- Poor initial estimates
  - Ill defined programs
- Different program than originally conceived
  - Different procurement quantities
  - Requirement changes
- Inefficiencies
  - Too many people
  - Too much money

- Lack of focus

- Other

(*Cost Growth of Major Defense Programs*, 2003:6)

The main objective of the study is to identify how much of cost growth is attributable to: 1) decisions: discretionary changes to the system relative to the description at milestone II , and 2) mistakes: changes not attributable to discretionary changes post milestone II. Also, a main objective is to establish a historical record for comparison between systems (*Cost Growth of Major Defense Programs*, 2003:10). The results of the study follow:

- Cost growth appears to have a correlation with commodity
- Cost estimating assumptions account for majority of mistakes cost growth
- Under estimating engineering effort is major source of RDT&E growth
- Nearly half of perceived cost growth is content change (i.e. decisions)
- Procurement cost growth is primarily due to optimistic learning curves
- Majority of systems do not have significant growth
- Higher cost systems appear to have less growth

(*Cost Growth of Major Defense Programs*, 2003:66).

Note that this study, like Sipple, Bielecki, and Moore's, evaluate cost growth as of the EMD phase of the system life cycle. Where this study differs is that the OSD and company do not focus on a single SAR cost variance category or a single appropriation. Instead, they seek to categorize cost growth into one of two categories: decisions or mistakes. From the results of their study we take away their finding that cost estimating assumptions account for the majority of cost growth in the mistakes category. This is consistent with most of our research as it reemphasizes the importance of generating accurate cost estimates up front and early in the acquisition process.

Chapter Summary

In this chapter, we discussed how the current acquisition process works as compared to how it used to work and explained why our study would need to analyze the old business practices. We also explored why accurate cost estimating is critical in today's acquisition environment, with heavy oversight, multiple stakeholders, scarce funding and numerous worldwide threats and ways to mitigate them. Upon examining the current acquisition environment we pointed out how risk is inherent in cost estimating due to countless unknowns, and that it is crucial to discover and address these unknowns up front and early. Finally we highlighted the relevant findings of recent studies in this area in order that we may approach our own research with an arsenal of "lessons learned".

III. Methodology

Chapter Overview

This chapter addresses the database to be used in our research, the response and predictor variables found in that database, the two-step methodology, important assumptions, and the methodology for integrating those variables into logistic and regression models to predict cost growth.

Database

We will use the same database originally created by Sipple and later updated by Bielecki and Moore. We do make several changes, however:

- We update the data with any new SARs (i.e. 2002). This entails adding four programs that meet the eligibility criteria (to be discussed later)
- We validate existing data with most current SAR and make necessary corrections.
- We review empty data fields and populate them (if the data can be found)
- We number the predictor variables for ease of use
- We delete, modify and add, several predictor variables (to be discussed later)

We accomplish these tasks using the same assumptions as Sipple: The program must be mature (i.e. be at least three years into the EMD phase) and use the Milestone I,II,III format (as opposed to the new A,B,C format). We also acknowledge that the SARs do have a number of limitations, originally pointed out in the 1992 RAND report by Paul Hough (Bielecki, 2003):

- Failure of some programs to use a consistent baseline cost estimate
- Exclusion of some significant elements of cost
- Exclusion of certain classes of major programs (e.g., special access programs)
- Constantly changing preparation guidelines
- Inconsistent interpretation of preparation guidelines across programs
- Unknown and variable funding levels for program risk
- Cost sharing in joint programs
- Reporting of effects of cost changes rather than their root causes

Despite these limitations, we conclude that the SAR is the best available source for this type of data. For a detailed discussion on these limitations, see Bielecki (2003:27-32). Please note that there are four variables that were populated from the Rand Report. These variables include *prototype, prototype phase, modification, weapon type, whether or not the program had a MS I,* and *service.* The complete database contains 135 data points (acquisition programs), four response variables, and 81 predictor variables from major DoD acquisition programs reported on from 1990-2002.

Response Variables

Our models have four different response variables, two discrete and two continuous. The discrete response variables predict whether a program will experience cost growth:

- *Engineering Cost Growth? Procurement* – binary variable: 1 for yes and 0 for no
- *Schedule Cost Growth? Procurement* – binary variable: 1 for yes and 0 for no

Our continuous response variables predict the amount of cost growth that will be experienced, in percentage format. We use the percentage format (instead of dollars) to achieve standardized responses from programs of different size. For reasons that will be explained later, this variable is transformed using a natural log.

- *Engineering Cost Variance%, Procurement* – continuous variable, percentage
- *Schedule Cost Variance%, Procurement* – continuous variable, percentage

Predictor Variable Updates

The predictor variables used by Sipple, Bielecki and Moore are quite extensive, and in many cases very complex. With the intent of scrubbing our data, we review all of the predictor variables to determine if any need to be removed, modified or added. First, we remove the following variables for the reasons stated.

- *Maturity from MSII in mos*
    - In cases where the latest SAR takes place after MSIII, the time between the two points is incorrectly added to the full length of the EMD phase
- *Actual Length of EMD using FUE-MSII in mos/FUE-based Maturity of EMD%*
    - FUE and IOC are interchangeable terms, and the database now contains an IOC variable that incorporates both
- *MSIII Complete*?
    - Users of our model would have no need to predict cost growth in the EMD phase if MSIII had already completed
- *RAND Concurrency Measurement Interval & RAND Concurrency Measurement Interval %*
    - These measurements assume MSIII has already occurred, in which case the user would have no need to predict EMD phase cost growth

- *Class at Least S*

    - Since we do not look at programs with security clearance higher than secret, and already have a variable for secret programs, this variable is redundant

- *Terminated?*

    - Users would have no need to predict cost growth if their program was terminated

- *Qty in PE*

    - Removed because it had only 7 observations

Next, we update the following variable to correct a flaw in the formula which allows impossible values to be calculated. We modify other variables in name only, to better reflect what the variable represents. These name changes are minor and are not reflected here.

- *Maturity of EMD %*

    - The way in which this variable was previously calculated, values of greater than 100% were obtained in many cases. Logic was built into the math which prevents this

Finally, we add the following variables to the list originally created by Sipple and updated by Bielecki and Moore.

- *ACAT 1*? – binary variable: 1 for yes and 0 for no
- *Service = Marines Only* – binary variable: 1 for yes and 0 for no
- *LRIP Qty Planned* – continuous variable to indicate the quantity in the baseline estimate
- *LRIP Qty Current Estimate* – continuous variable to indicate the quantity as currently estimated in the latest SAR

- *LRIP Planned?* – binary variable: 1 for yes and 0 for no; indicates if the program had LRIP planned

- *% R&D of Total Program* – continuous variable calculated by dividing *Length of R&D in Funding Yrs* by *Funding YR Total Program Length*

- *% Prod of Total Program* – continuous variable calculated by dividing *Length of Prod in Funding Yrs* by *Funding YR Total Program Length*

- *Fund Years of R&D + Proc Complete* – continuous variable calculated by adding Funding *Years of R&D Completed* and *Funding Years of Proc Completed*

- *Length of R&D + Proc Funding Years* – continuous variable calculated by adding *Length of R&D in Funding Years* and Length of *Proc in Funding Years*

Predictor Variables

The following pages reflect the aforementioned deletions, changes and additions, to the predictor variables. We continue to use Sipple's variable categories: program size, physical type of program, management characteristics, schedule characteristics, and other characteristics. This is the complete list of predictor variables we use in our logistic and multiple regression models. For the purpose of a clean presentation, we do not use the numbering scheme here, but it is reflected in the Appendix.

<u>Program Size Variables</u>

- *Total Cost CY $M 2002* – continuous variable which indicates the total cost of the program in CY $M 2002

- *Total Quantity* – continuous variable which indicates the total quantity of the program at the time of the SAR date; if no quantity is specified, we assume a quantity of one (or another appropriate number) unless the program was terminated

- *Unit Cost* – continuous variable that equals the quotient of the total cost and total quantity variables above

- *Qty Planned for R&D* – continuous variable which indicates the quantity in the baseline estimate

- *Qty Currently Estimated for R&D* – continuous variable that indicates the quantity that was estimated in the Planning Estimate

- *ACAT 1*? –binary variable: 1 for yes and 0 for no

- *LRIP Qty Planned* – continuous variable to indicate the quantity in the baseline estimate

- *LRIP Qty Current Estimate* – continuous variable to indicate the quantity as currently estimated in the latest SAR

- *LRIP Planned?* – binary variable: 1 for yes and 0 for no; indicates if the program had LRIP planned

<u>Physical Type of Program</u>

- Domain of Operation Variables

    - *Air* – binary variable: 1 for yes and 0 for no; includes programs that primarily operate in the air; includes air-launched tactical missiles and strategic ground-launched or ship-launched missiles

    - *Land* – binary variable: 1 for yes and 0 for no; includes tactical ground-launched missiles; does not include strategic ground-launched missiles

    - *Space* – binary variable: 1 for yes and 0 for no; includes satellite programs and launch vehicle programs

    - *Sea* – binary variable: 1 for yes and 0 for no; includes ships and ship-borne systems other than aircraft and strategic missiles

- Function Variables

    - *Electronic* – binary variable: 1 for yes and 0 for no; includes all computer programs, communication programs, electronic warfare programs that do not fit into the other categories

    - *Helo* – binary variable: 1 for yes and 0 for no; helicopters; includes V-22 Osprey

    - *Missile* – binary variable: 1 for yes and 0 for no; includes all missiles

    - *Aircraft* – binary variable: 1 for yes and 0 for no; does not include helicopters

    - *Munition* – binary variable: 1 for yes and 0 for no

    - *Land Vehicle* – binary variable: 1 for yes and 0 for no

    - *Space (Rand)* –binary variable: 1 for yes and 0 for no

    - *Ship* – binary variable: 1 for yes and 0 for no; includes all watercraft

    - *Other* – binary variable: 1 for yes and 0 for no; any program that does not fit into one of the other function variables

<u>Management Characteristics</u>

- Military Service Management

    - *Svs > 1* – binary variable: 1 for yes and 0 for no; number of services involved at the date of the SAR

    - *Svs > 2* – binary variable: 1 for yes and 0 for no; number of services involved at the date of the SAR

    - *Svs > 3* – binary variable: 1 for yes and 0 for no; number of services involved at the date of the SAR

    - *Service = Navy Only* – binary variable: 1 for yes and 0 for no

    - *Service = Joint* – binary variable: 1 for yes and 0 for no

    - *Service = Army Only* – binary variable: 1 for yes and 0 for no

    - *Service = Marines Only* – binary variable: 1 for yes and 0 for no

    - *Service = AF Only* – binary variable: 1 for yes and 0 for no

    - *Lead Svc = Army* – binary variable: 1 for yes and 0 for no

    - *Lead Svc = Navy* – binary variable: 1 for yes and 0 for no

    - *Lead Svc = DoD* – binary variable: 1 for yes and 0 for no

- *Lead Svc = AF* – binary variable: 1 for yes and 0 for no
- *AF Involvement* – binary variable: 1 for yes and 0 for no
- *N Involvement* – binary variable: 1 for yes and 0 for no
- *MC Involvement* – binary variable: 1 for yes and 0 for no
- *AR Involvement* – binary variable: 1 for yes and 0 for no

o Contractor Characteristics

- *Lockheed-Martin* – binary variable: 1 for yes and 0 for no
- *Northrup Grumman* – binary variable: 1 for yes and 0 for no
- *Boeing* – binary variable: 1 for yes and 0 for no
- *Raytheon* – binary variable: 1 for yes and 0 for no
- *Litton* – binary variable: 1 for yes and 0 for no
- *General Dynamics* – binary variable: 1 for yes and 0 for no
- *No Major Defense Contractor* – binary variable: 1 for yes and 0 for no; a program that does not use one of the contractors mentioned immediately above = 1
- *More than 1 Major Defense Contractor* – binary variable: 1 for yes and 0 for no; a program that includes more than one of the contractors listed above = 1
- *Fixed-Price EMD Contract* – binary variable: 1 for yes and 0 for no

<u>Schedule Characteristics</u>

o RDT&E and Procurement Maturity Measures

- *Maturity (Funding Yrs complete)* – continuous variable which indicates the total number of years completed for which the program had RDT&E or procurement funding budgeted

- *Funding YR Total Program Length* – continuous variable which indicates the total number of years for which the program has either RDT&E funding or procurement funding budgeted

- *Funding Yrs of R&D Completed* – continuous variable which indicates the number of years completed for which the program had RDT&E funding budgeted

- *Funding Yrs of Proc Completed* – continuous variable which indicates the number of years completed for which the program had procurement funding budgeted

- *Length of Proc in Funding Yrs* – continuous variable which indicates the number of years for which the program has procurement funding budgeted

- *Length of R&D in Funding Yrs* – continuous variable which indicates the number of years for which the program has RDT&E funding budgeted

- *R&D Funding Yr Maturity %* – continuous variable which equals *49 Funding Yrs of R&D Completed* divided by *52 Length of R&D in Funding Yrs*

- *Proc Funding Yr Maturity %* – continuous variable which equals *50 Funding Yrs of Prod Completed* divided by *51 Length of Prod in Funding Yrs*

- *Total Funding Yr Maturity %* – continuous variable which equals *Maturity (47 Funding Yrs complete)* divided by 48 *Funding YR Total Program Length*

- *Fund Years of R&D + Prod Complete* – continuous variable calculated by adding Funding *Years of R&D Completed* and *Funding Years of Prod Completed*

- *Length of R&D + Prod Funding Years* – continuous variable calculated by adding *Length of R&D in Funding Years* and Length of *Prod in Funding Years*

- EMD Maturity Measures

  - *Actual Length of EMD* – continuous variable calculated by subtracting the earliest MS II date from the latest MS III date indicated

- *Maturity of EMD %* – continuous variable calculated by dividing *Maturity from MS II (current calculation in months)* by *56 Actual Length of EMD*

- *Time From MSII to IOC in months* – continuous variable calculated by subtracting the earliest MS II date from the IOC date

- *Maturity of EMD at IOC %* – continuous variable calculated by dividing *Maturity from MS II (current calculation in months)* by *57 Time From MSII to IOC in months*

- Concurrency Indicators

  - *Proc Started based on Funding Yrs* – binary variable: 1 for yes and 0 for no; if procurement funding is budgeted in the year of the SAR or before, then = 1

  - *Proc Funding before MS III* – binary variable: 1 for yes and 0 for no

<u>Other Characteristics</u>

- *# Product Variants in this SAR* – continuous variable which indicates the number of versions included in the EMD effort that the current SAR addresses

- *Class – S* – binary variable: 1 for yes and 0 for no; security classification Secret

- *Class – C* – binary variable: 1 for yes and 0 for no; security classification Confidential

- *Class – U* – binary variable: 1 for yes and 0 for no; security classification Unclassified

- *Risk Mitigation* – binary variable: 1 for yes and 0 for no; indicates whether there was a version previous to SAR or significant pre-EMD activities

- *Versions Previous to SAR* – binary variable: 1 for yes and 0 for no; indicates whether there was a significant, relevant effort prior to the DE; a pre-EMD prototype or a previous version of the system would apply

- *Modification* – binary variable: 1 for yes and 0 for no; indicates whether the program is a modification of a previous program

- *Prototype* – binary variable: 1 for yes and 0 for no; indicates whether the program had a prototyping effort

- *Dem/Val Prototype* – binary variable: 1 for yes and 0 for no; indicates whether the prototyping effort occurred in the PDRR phase

- *EMD Prototype* – binary variable: 1 for yes and 0 for no; indicates whether the prototyping effort occurred in the EMD phase

- *PE?* – binary variable: 1 for yes and 0 for no; indicates whether the program had a Planning Estimate

- *Significant pre-EMD activity immediately prior to current version* – binary variable: 1 for yes and 0 for no; indicates whether the program had activities in the schedule at least six months prior to MSII decision

- *Program have a MS I?* – binary variable: 1 for yes and 0 for no

- *% R&D of Total Program* – continuous variable calculated by dividing *Length of R&D in Funding Yrs* by *Funding YR Total Program Length*

- *% Proc of Total Program* – continuous variable calculated by dividing *Length of Proc in Funding Yrs* by *Funding YR Total Program Length*

Model Building

The only assumption about the regression model itself is that the response variables are reasonably continuous. The distributions of the response variables for Sipple (2002), Bielecki (2003) and Moore (2003) were mixed (having both continuous and discrete characteristics). This is the result of a mass at 0 (no cost growth). The consequences of building multiple regression models that violate the assumption of being reasonably continuous translate into a violation of the assumption of homoscedasticity of the residuals. As can be seen in Figure 3.1, the distributions of our two continuous response variables are also mixed, with a large mass at 0:

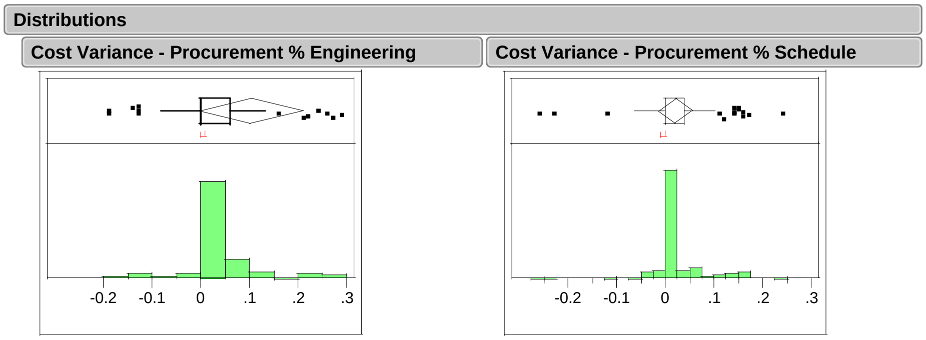


Figure 3.1 - Histograms of Engineering and Schedule responses

Given past research and our obviously mixed distribution, we will also employ the 2-step methodology of using logistic regression to predict whether there will be cost growth, and multiple linear regression to predict how much that cost growth will be. But before we can start building models, we partition the database.

<u>Partitioning of the Database</u>

Validation set partitioning must be done prior to model building, otherwise we will not be able to validate the models that we create. "Validation" is essentially a test to see how well our models predict, using data points that we did not use in the building of the model.

First, we reorder the data points (programs) in the database using a random number generator. This is to ensure there is no bias between which data points are used to build and validate. Then, we create a subset using the first 108 programs (this is 80% of the database). This 80% is be used in all model building for the duration of our research. Then, subset is created with the remaining 27 data points (20%), and saved for validation.

Logistic Regression

By design, the purpose of a logistic regression model is to make a binomial “yes or no” prediction. “1” means yes, and “0” means no. For our research, we will build two logistic regression models to predict whether cost growth will occur in the *engineering* and *schedule* categories of programs funded with procurement dollars in the EMD phase.

Given that we have 81 predictor variables, there are more than 30 billion possible combinations of 8-variable models. While the only way to ensure we find the best model (8 variable or less) is to run all the combinations in statistical software program, it would be too time consuming to finish. For this reason, we develop a sound methodology to hunt for highly predictive models.

First, we run all the variables by themselves and record their pertinent statistics. This includes the p-values of the individual parameters, the average p-value, the $R^2(U)$ score, and the receiver operator characteristic (ROC) score. We explain these evaluation measurements in chapter 4. From these models, we choose the best 10 to 15 models, looking primarily for individual p-values of 0.05 or less. These are our 1-variable models.

We then run each of these best 1-variable models against all other variables to find combinations that perform well together, using the aforementioned criteria. These results are recorded. From this analysis, we find our best 2-variable models. Then we look at these models to see if any become “marginal” when the second variable is added. Any “marginal” models will be removed. Once we have our 2-variable models we look at the p-values of the individual parameters of all models created (both rounds) and look for variables whose p-value did not score below 0.1 either by themselves, or in combination with any other variable. Any variable that meets this criteria will be excluded from future models. We do this to reduce the total amount of variables that have to be run, with the assumption that any variable that does not perform well by itself, or with any other variable, will not make it into any final model.

We add the remaining variables in the same manner as the second variable, minus the "exclusion" step. In other words, all variables that make it past the 2-variable round will be tested in all remaining rounds. We will stop adding variables to the model when the data point-to-variable ratio drops under 10, or adding more variables no longer yields favorable results, whichever comes first. As our data point-to-variable ratio drops below 10, we begin to run the risk of over fitting our model, which reduces is statistical validity.

The previous steps give us our final full model. This model contains the "core" variables that will be in our final reduced model. We find our final reduced model by manipulating the final full model. This "manipulation" includes, but is not limited to, discretizing continuous variables, adding interaction terms, removing variables and mathematically combining variables. Once we have our final reduced model, we will run it with the variables we excluded after the second round to make sure they do not contribute in a positive way. Finally, we validate the final reduced model using the saved 27 data points.

Multiple Linear Regression

Multiple linear regression, also known as ordinary least squares (OLS) regression, is used to predict the level of positive cost growth. We have two reasons for predicting only positive cost growth. The first is that we know we cannot use the range of positive to negative cost growth because it violates the assumption of being reasonably continuous (i.e. it contains a discrete mass at zero). The second is that the primary concern of the user is positive cost growth (i.e. no or negative cost growth is good).

Another lesson that we have learned from past research in this area is that the assumption of constant variance of the residuals always fails unless you transform them using a natural log. Figure 3.2 is a histogram of only the positive growth responses, before the transformation:

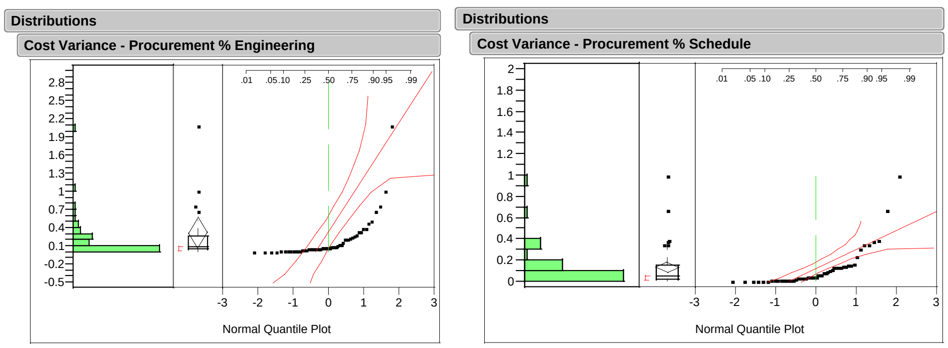


Figure 3.2 - Histograms of Only Positive Responses

Given past research, we expect a log-normal distribution, and that is exactly what we get. From this we make the assumption, that like past models dealing with cost growth, we must take the natural log of the responses to avoid heteroscedasticity of the model's residuals. When we do, we get somewhat normal distributions, as seen in Figure 3.3.

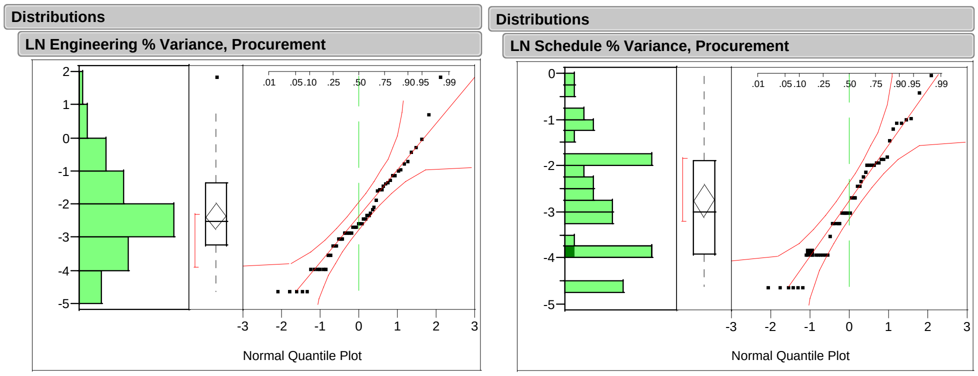


Figure 3.3 - Histograms of Only Positive Responses (transformed)

Now that we have transformed response variables, we build our models in the same manner we build our logistic regression models. The only differences being that the measurement of adjusted R^2 replaces the measurement of R^2(U), and we do not have a ROC score.

Chapter Summary

This chapter details the overall methodology for building our predictive models. We discuss how we update the database which includes deletions, modifications and additions. We also provide a current and comprehensive list of response and predictor variables that we use in model building, as well as the methodology for generating combinations of variables with predictive ability. We present our findings in the following chapter.

IV. Analysis Results

Chapter Overview

This chapter describes the results of our two logistic and multiple regression model analyses. First, we conduct a preliminary data analysis on the database. Second, we discuss our evaluation measures for logistic regression and our logistic regression results. Finally, we discuss our evaluation measurements for OLS regression and our OLS regression results.

Preliminary Data Analysis

As previously stated, the purpose of this analysis is to determine what, if any, factors in the DoD acquisition community can be identified to accurately predict cost growth such that the risk of cost growth from the baseline estimate is reduced. We use a two-pronged approach which first predicts whether positive cost growth will occur (using logistic regression), then the extent to which it will occur (using OLS regression). The scope of our study includes the Engineering and Schedule SAR cost categories, with the response variables being cost growth in the EMD phase. We only use 1990-2002 SAR data that have a DE baseline and a procurement appropriation.

As stated in Chapter Three, the only assumption about the regression model itself is that the response variables are reasonably continuous. The distributions of the response variables for Sipple (2002), Bielecki (2003) and Moore (2003) were mixed (having both continuous and discrete characteristics). This is the result of a mass at 0 (no cost growth). The consequences of building multiple regression models that violate the assumption of being reasonably continuous translate into a violation of the assumption of homoscedasticity of the residuals. As reiterated in Figure 4.1, the distributions of our two continuous response variables are also mixed, with a large mass at 0. For this reason, we also employ the two-step process for each response variable by first performing logistic regression, then concluding with OLS.

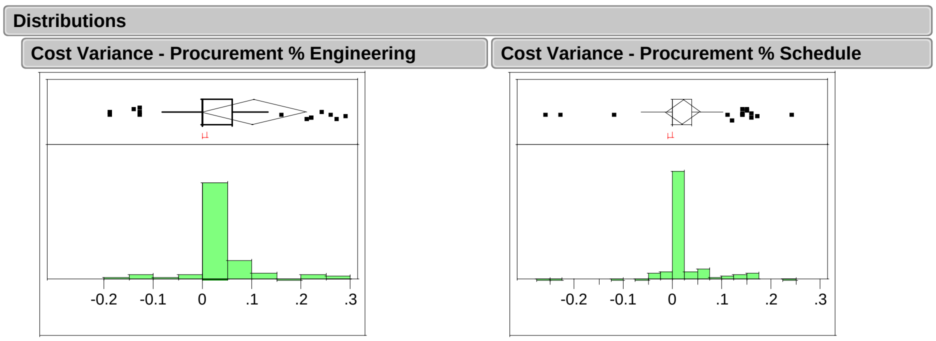
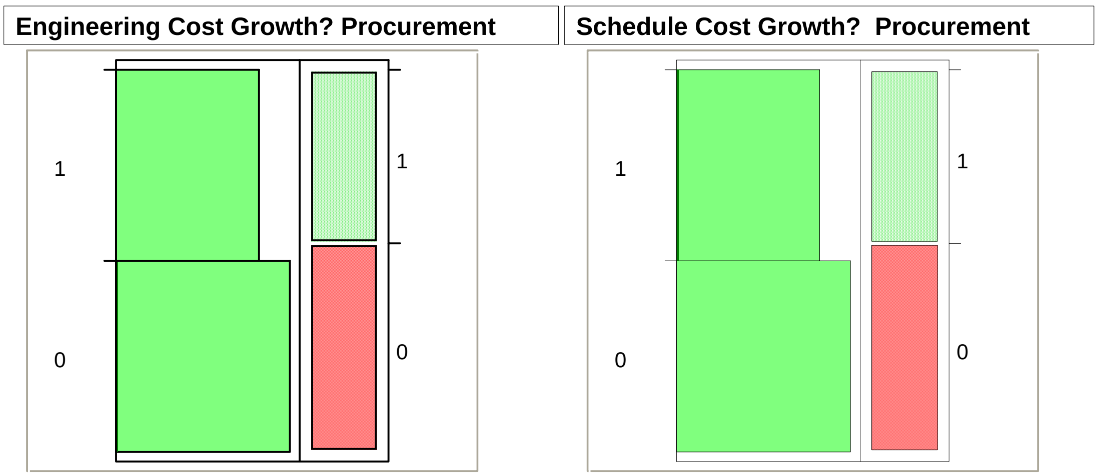


Figure 4.1 - Histograms of All Responses

The second issue we address during the preliminary data analysis is whether the response variables will be conducive to building and validating regression models. Specifically, we want to know two things. First, will there be enough data points to both build the model and validate it? Second, is the range of responses such that relationships can be discerned? For example, if all the programs had the same responses, we would not be able to discern relationships given differences in the independent variables. To accomplish this, we look at the distributions of the various responses.

Figure 4.2 shows the distributions of the binomial cost growth responses for the Engineering and Schedule categories. We see that both distributions contain 135 total observations. This means that we will have 108 data points from which to build logistic regression models (80%) and 27 data points for validation (20%). If we apply a strict 10:1 data point to variable ratio, our models could contain up to 10 variables, so we conclude that there are plenty of data points. We also see that the distributions contain a desirable variety in the responses, as both contain 74 “trues” and 61 “falses”. We conclude that the database is suitable for developing logistic regression models.



Frequencies (Engineering Cost Growth? Procurement)

Level	Count	Prob
0	74	0.54815
1	61	0.45185
Total	135	1.00000

2 Levels

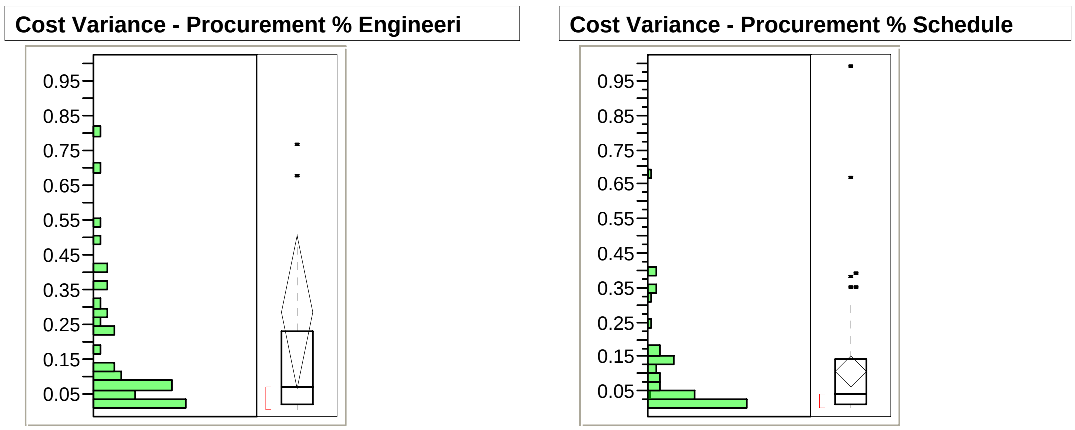
Frequencies (Schedule Cost Growth? Procurement)

Level	Count	Prob
0	74	0.54815
1	61	0.45185
Total	135	1.00000

2 Levels

Figure 4.2 - Distribution of Logistic Regression Responses

Figure 4.3 shows the distributions of the continuous cost growth responses for the Engineering and Schedule categories. We see that both distributions contain 61 total observations. This means that we will have 49 data points from which to build OLS regression models (80%) and 12 data points for validation (20%). If we apply a strict 10:1 data point to variable ratio, our models could contain up to five variables. While we would prefer the flexibility to have more, five is sufficient. We also see that the distributions, which visually appear to be log-normal, contain a wide variety of responses between 0 and 1. We conclude that the database is also suitable for developing OLS regression models.



Moments

Mean	0.2843852
Std Dev	0.866045
Std Err Mean	0.1108857
upper 95% Mean	0.5061897
lower 95% Mean	0.0625808
N	61

Moments

Mean	0.1064115
Std Dev	0.1698055
Std Err Mean	0.0217414
upper 95% Mean	0.1499007
lower 95% Mean	0.0629223
N	61

Figure 4.3 - Distribution of OLS Responses

Finally, the log-normal distributions in Figure 4.3 reiterate our discussion about the constant variance problems that Sipple encountered in his study. As stated in Chapter Three, Sipple had to transform the response variables in his OLS models using the natural log, in order to eliminate the violation of the assumption of constant variance. We will also use the natural log transformation, which gives us the more normal looking distributions in Figure 4.4.

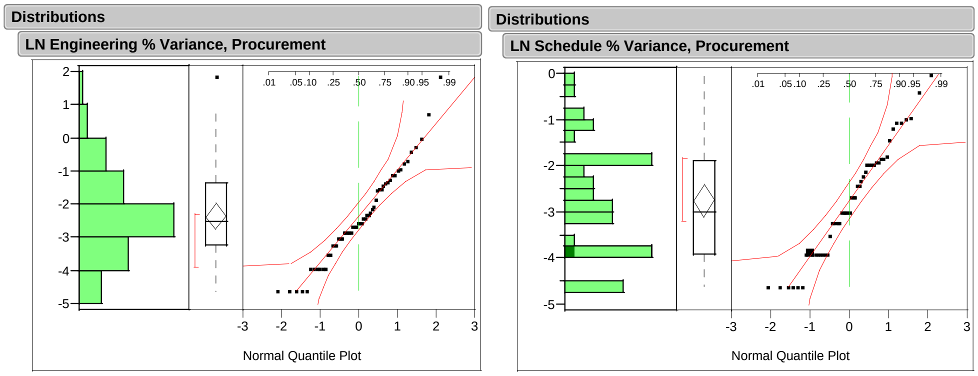


Figure 4.4 - Transformed Distribution of OLS Responses

Evaluation Measures for Logistic Regression

We use four evaluation measures during our model building process: Average P-Values, $R^2(U)$, Area Under the Receiver Operator Characteristic (ROC) Curve, and the Data Point-to-Variable Ratio. We explain each of these evaluation measures in the following paragraphs.

While there is a whole-model p-value calculation made in JMP®, we will not use it because it does not effectively assist us in differentiating between models (it is always very low). Instead, we look at the p-values of the individual variables and their estimated parameters. P-values tell us if there is statistical significance between an individual variable and the response variable, given its parameter estimate and the other variables in the model. P-values lower than 0.05 indicate statistical significance with a 95% level of confidence. For the purpose of measuring the overall significance of the model, we take the sum of the individual p-values and divide them by the number of variables in the model to arrive at an average p-value. Lower average p-values will mean more predictive models.

According to Moore, “The R^2(U) that JMP® uses is the difference of the negative log likelihood of the fitted model minus the negative log likelihood of the reduced model divided by the negative log likelihood of the reduced model (Moore, 2003:33)”. In other words, it “is the proportion of the total uncertainty that is attributed to the model fit (JMP® 5.1, 2003:Help)”. In laymen’s terms, R^2(U) is a number between 0 and 1 that tells us how well our model fits our data; 0 being not at all, and 1 being perfectly.

To understand what a ROC curve is, we must first define two terms: sensitivity and specificity. Sensitivity is the proportion of true positives that our model produces, out of all possible positives. Specificity is the proportion of true negatives that our model produces, out of all possible negatives. The area under the ROC curve is discovered by plotting sensitivity (proportion of true positives) against 1-specificity (proportion of false positives). This gives us a number between 0 and 1, with the goal to get as much area under the curve as possible (i.e. closer to 1).

Finally, the data point-to-variable ratio is the number of observations our model is able to use (given the variables selected) divided by the number of predictor variables in the model. A good rule of thumb is to have a ratio of no less than 10:1, but according to Neter et al., we could go as low as 6:1 (Neter, 1996:437). The lower the ratio gets, the better the chance that our model will become statistically invalid. For more information on the data point-to-variable ratio, see Sipple (2002).

Logistic Regression Results

Logistic regression is an effective tool for predicting binary outcomes. The response variable for our logistic regression model is a Bernoulli random variable. The two possible outcomes for this variable are 0 (no positive cost growth) and 1 (positive cost growth). Therefore, our models will attempt to predict whether cost growth will occur. We build two logistic regression models, one to predict whether cost growth will occur in the Engineering cost category, and the other to predict whether cost growth will occur in the Schedule cost category.

We use JMP 5.1® statistical analysis software to build our models using the methodology described in Chapter Three. The following pages contain the results for the Engineering and Schedule logistic regression models.

Engineering Cost Category

The first model we build uses the Engineering Cost Growth response variable. We begin by regressing each predictor variable against the response variable and employing the evaluation measurements. In round one, we are mostly concerned with finding variables whose p-values are less than 0.05. Our regressions yield a total of 15 of these variables, which will be the "foundation" for further regressions.

In round two, we take each of the 15 variables from round one, and run each of them with all other variables to find the best 15 two-variable models. At this point we begin looking for the lowest average p-value to determine which variable performs the best with each of our original 15 foundation variables. If a variable does not achieve an average p-value of 0.05 or less with any other variable, it is eliminated as a foundation variable. Also, any models that contain individual variables whose p-values exceed 0.05 are also not considered. In round two, three of the original 15 variables fail to achieve an average p-value of 0.05 or less, leaving us with 12 two-variable models.

Before continuing on to round three, we review how well our 81 predictor variables have performed up to this point. We discover that 51 of the 81 predictor variables do not achieve an individual p-value of 0.05 or less in round one, nor do they achieve an average p-value of 0.05 or less with any other variable. From this we conclude that these variables have very little chance of making it into any final model since they don't perform well by themselves or with any other variable. In the interest of reducing the number of models we have to run, we exclude these variables in the following rounds, but will run them with our best models towards the end to confirm that they offer no additional predictive ability.

In round three, we run our 12 two-variable models with each of the 30 un-excluded variables. At this point we begin to look at evaluation measurements more closely. First, we check to see if the model has any individual p-values above 0.05. If it does, we don't consider it unless they are marginal (i.e. close to 0.05). Next, we look at number of observations. Models that violate the 10:1 data point to variable ratio are not considered. Third, we look at average p-values. Fourth, we look at $R^2(U)$. Finally, we look at the ROC score.

Deciding whether a variable will be added to the two-variable model is a combination of the last three evaluation measures. In most cases, the models with the lowest average p-values also yield the highest $R^2(U)$ and ROC scores, and are therefore easy to pick. In the rare instance where this is not the case, we make a judgment call which results in one of two outcomes: picking one model over another, or carrying both models forward to the next round. We repeat this process until the addition of variables no longer yields favorable results. When we reach this point, we have 48 models ranging in size from two to seven variables.

We choose to make model 48 our final full model because it has the best R^2(U) and ROC scores, and still has an average p-value lower than 0.05. As promised, we run the final full model with all the variables excluded after round two, and confirm that they offer no further improvement. Table 4.1 summarizes the evaluation measurements for the best model of each size (from two to seven variables). We do not include the data point-to-variable ratio for these models, because they all use at least 100 observations, easily meeting our data point-to-variable ratio criteria. Attachment 6 provides the full JMP® analysis of model 48.

Table 4.1 - Best 2-7 Variable Logistic Regression Models (Engineering)

Best Models Using 2-7 Variables				
Model#	**# Variables**	**Ave P-Value**	**R^2(U)**	**ROC**
2	2	0.0017	0.1517	0.7474
14	3	0.0105	0.2019	0.7758
27	4	0.0131	0.2359	0.8094
38	5	0.0111	0.2620	0.8292
45	6	0.0243	0.2760	0.8348
48	7	0.0346	0.2992	0.8562

Now that we have chosen our final full model, we perform the manipulations mentioned in Chapter Three to see if we can make a better, reduced model. This yields us one promising observation: Discretizing variable 54 by making any value above 0.45 true, we notice a significant improvement in its individual p-value. Discretizing is the term we use to describe the process of transforming a continuous variable into a discrete variable. The only problem is that this manipulation, while improving the overall model, causes the individual p-value of variable 12 to increase beyond an acceptable level. We then remove variable 12, and discover a six variable model that meets our criteria. No further manipulations to the reduced model yield favorable results. Table 4.2 compares the final full model to this reduced candidate.

Table 4.2 - Full/Reduced Model Comparison (Engineering)

Full/Reduced Model Comparison				
Model#	**# Variables**	**Ave P-Value**	**R^2(U)**	**ROC**
48	7	0.0346	0.2992	0.8348
49	6	0.0219	0.3264	0.8622

Compared to the final full model, 49 has the highest R^2(U), and the highest ROC score. It also yields lower average p-values, and a data point-to-variable ratio in excess of 16:1. Given these evaluation measures, we choose model 49 as our final reduced model. To see all 49 models and their variables, refer to Attachments 1 and 5. Attachment 7 provides the full JMP® analysis on model 49. The predictor variables used in model 49 are:

Length of R&D in Funding Years
Classification Secret?
LRIP Planned?
Lockheed Martin?
Discretized Variable 54 (Proc Funding Yr Maturity % >.45?)
Risk Mitigation?

The final step for our reduced model is to validate it using the 20% of the database that we set aside prior to model building. We do this by putting the entire database back together, running our reduced model, and comparing the actual response to the most likely response (generated by our model).

Of the 27 data points we use in the validation, 26 are usable (96.3%). Out of all 135 data points in the database, 126 are usable (93.3%). From this we conclude that our model is highly universal. Of the 26 observations we use in the validation, our model yields 13 accurate predictions (50%). Out of all 126 observations, our model yields 94 accurate predictions (74.6%). Given this disparity between the full database and the validation set, we explore the potential differences between the 20% validation set and the 80% database. We do this by looking at the distributions of each variable in both the 20% and 80% database. By randomizing the partitioning process in the beginning, we hope to get an even distribution of data amongst our predictor variables. What we discover when we compare these distributions, is that we did not. Table 4.3 compares the mean values for each variable in both databases.

Table 4.3 - Comparison of Variable Means for the 20% and 80% Databases

Variable Means for Both Databases			
Variable	**20%**	**80%**	**% Difference**
52	13.31	15.10	-11.9%
65	0.29	0.34	-14.7%
77	0.37	0.42	-11.9%
38	0.18	0.25	-28.0%
82	0.48	0.44	9.1%
68	0.81	0.80	1.3%

Note that the mean for four of the six variables is noticeably lower (more than 10% difference) in the 20% database than in the 80% database. Variable 38 has a substantial 28% lower mean in the 20% database. We conclude from this that that our randomly selected 20% database does not accurately represent the entire database as it will yield noticeably lower values, and offers at least a partial explanation for our low validation score. The JMP Analysis for these distributions can be found in attachments 10 and 11.

In conclusion, our logistic regression model to predict cost growth does have predictive capability as it does accurately predict cost growth for the programs in our

100% database 74.6% of the time, despite a poor validation score. We also conclude that it is highly universal, as 93.3% of all the programs studied were usable.

Schedule Cost Category

The second model we build uses the Schedule Cost Growth response variable. In round one, 12 variables have p-values of less than 0.05. In round two, we discover that three foundation variables work very well with two other variables, so both models (for each of these foundation variables) will be used in the next round, leaving us with 15 two-variable models. Of our 81 predictor variables, we discover that 60 do not achieve an individual p-value of 0.05 or less in round one, nor do they achieve an average p-value of 0.05 or less with any other variable. From this, we conclude that these variables have very little chance of making it into any final model since they don't perform well by themselves or with any other variable, so they are excluded. Continuing the process from round three on, we finish with 35 models ranging in size from two to five variables.

We choose to make model 28 our final full model because it has the best R^2(U) and ROC scores, and still has an average p-value lower than 0.05. As promised, we run the final full model with all the variables excluded after round two, and confirm that they offer no further improvement. Table 4.4 summarizes the evaluation measurements for the best model of each size (from two to four variables). We do not include the data point-to-variable ratio for these models, because they all use at least 100 observations, easily meeting our data point-to-variable ratio criteria. Attachment 8 provides the full JMP® analysis of model 48.

Table 4.4 - Best 2-4 Variable Logistic Regression Models (Schedule)

Best Models Using 2-4 Variables				
Model#	**# Variables**	**Ave P-Value**	**R^2(U)**	**ROC**
1	2	0.0006	0.2696	0.8356
16	3	0.0099	0.3062	0.8478
28	4	0.0163	0.3375	0.8667

Manipulation of model 28 yield us favorable results. First, when we discretize variable 54 such that all values above 0.43 are true, our evaluation measurements improve. Second, when we add an interaction term between variables 36 and 77, our measurements improve further. Table 4.5 compares the final full model to this reduced candidate.

Table 4.5 - Full/Reduced Model Comparison (Schedule)

Full/Reduced Model Comparison				
Model#	**# Variables**	**Ave P-Value**	**R^2(U)**	**ROC**
28	4	0.0163	0.3375	0.8667
36	5	0.0172	0.4392	0.8913

Compared to the final full model, 36 has the highest R^2(U), and the highest ROC score. Its average p-values are slightly higher, but still well below 0.05, and has a data point-to-variable ratio in excess of 20:1. Given these evaluation measures, we choose model 36 as our final reduced model. To see all 36 models and their variables, refer to Attachments 2 and 5. Attachment 9 provides the full JMP® analysis on model 36. The predictor variables used in model 36 are:

Discretized Variable 54 (Proc Funding Yr Maturity % >.43?)
68 Risk Mitigation?
Marine Core Involvement?
LRIP Planned?
Interaction Term (MC Involvement? and LRIP Planned?)

Next, we validate our model. Of the 27 data points we use in the validation, 26 are usable (96.3%). Out of all 135 data points in the database, 128 are usable (94.8%). From this, we conclude that our model is highly universal. Of the 26 observations we use in the validation, our model yields 13 accurate predictions (50%). Out of all 128 observations, our model yields 94 accurate predictions (78.1%). Given this disparity between the full database and the validation set, we must again explore the potential differences between the 20% validation set and the 80% database. We do this by looking at the distributions of each variable in both the 20% and 80% database. Table 4.6 compares the mean values for each variable in both databases.

Table 4.6 - Comparison of Variable Means for the 20% and 80% Databases

Variable Means for Both Databases			
Variable	**20%**	**80%**	**% Difference**
82	0.52	0.46	12.0%
68	0.81	0.80	0.5%
36	0.19	0.21	-13.1%
77	0.37	0.43	-13.1%

Note that the mean for three of the four variables is noticeably higher or lower (more than 10% difference) in the 20% database than in the 80% database. While we conclude from this that that our randomly selected 20% database does not accurately represent the entire database, it is does not have any substantial differences and thus offers little explanation for our poor validation results. The JMP® Analysis for these distributions can be found in attachments 12 and 13.

In conclusion, our logistic regression model to predict cost growth does have predictive capability as it does accurately predict cost growth for the programs in our 100% database 78.1% of the time, despite a poor validation score. We also conclude that it is highly universal, as 94.8% of all the programs studied are usable.

Evaluation Measures for OLS Regression

We use three evaluation measures during our model building process: Average P-Values, Adjusted R^2, and the Data Point-to-Variable Ratio. Save for replacing R^2(U) with Adjusted R^2, these evaluation measures are the same OLS as they are for logistic regression.

According to JMP 5.1®, R^2 "estimates the proportion of the variation in the response around the mean that can be attributed to terms in the model rather than to random error," and the adjusted R^2 "adjusts R^2 to make it more comparable over models with different numbers of parameters by using the degrees of freedom in its computation (JMP® 5.1, 2003:Help)". Given that we are comparing models with different numbers of parameters, we use adjusted R^2, instead of R^2. Adjusted R^2 is a number from 0 to 1, 0 being that none of the variation in the response can be attributed to the model, and 1 being that all of it can. In laymen's terms, the higher the adjusted R^2, the more predictive the model.

Another difference between OLS and logistic regression is that in OLS, we test assumptions after we select a final reduced model. We have three assumptions tests to conduct: Normality of the residuals, independence, and homoscedasticity of the residuals. We test for normality of the residuals by conducting a Shapiro-Wilkes analysis, which fits the studentized residuals to a normal distribution and calculates a p-value indicating how well they fit. P-values below .05 result in failure of this assumption. Since this is not time series data, we test for independence by looking at the various data points and discussing relationships they have to each other and how they were collected. We test homoscedasticity of the residuals by conducting a Breusch-Pagan calculation which results in a p-value. P-values below 0.05 indicate a failure of this assumption.

Once we have completed our assumption tests, we perform a Cook's Distance analysis. Cook's Distance is a tool that measures how far each data point used in the model is from the mean of that independent variable. Data points that are far away from the mean may have too much influence on the model, and will be a candidate for removal. Values from 0 to 0.25 are normal, values from 0.25 to 0.5 are marginal, and values above 0.5 will probably need to be removed. After Cook's Distance, we check the Variance Inflation Factor (VIF) scores. VIF scores check for linear dependency amongst predictors. This dependency may exist if VIF scores exceed 10. Models with linear dependency may run the risk of being statistically invalid.

OLS Regression Results

OLS regression is an effective tool for predicting reasonably continuous outcomes. In this case, we use OLS regression to predict the amount of cost growth that will occur in a program, once our logistic regression model predicts that cost growth will occur. We build two OLS regression models, one to predict the amount of cost growth that will occur in the Engineering cost category, and the other to predict the amount of cost growth will occur in the Schedule cost category. The following pages contain the results for the Engineering and Schedule OLS regression models.

Engineering Cost Category

The first model we build uses the natural log of the Engineering Cost Growth % response variable. We begin by regressing each predictor variable against the response variable and employing the evaluation measurements. In round one, we are mostly concerned with finding variables whose p-values are less than 0.05. Our regressions yield a total of five of these variables, which will be the "foundation" for further regressions.

In round two, we take each of the five variables from round one, and run each of them with all other variables to find the best five two-variable models. At this point we begin looking for the lowest average p-value to determine which variable performs the best with each of our original five foundation variables. If a variable does not achieve an average p-value of 0.05 or less with any other variable, it is eliminated as a foundation variable. Any models that contain individual variables whose p-values exceed 0.05 are not considered either.

Before continuing on to round three, we review how well our 81 predictor variables have performed up to this point. We discover that 73 of the 81 predictor variables do not achieve an individual p-value of 0.05 or less in round one, nor do they achieve an average p-value of 0.05 or less with any other variable. From this, we conclude that these variables have very little chance of making it into any final model since they don't perform well by themselves or with any other variable. In the interest of reducing the number of models we have to run, we exclude these variables in the following rounds, but will run them with our best models towards the end to confirm that they offer no additional predictive ability.

In round three, we run our five two-variable models with each of the eight un-excluded variables. At this point we begin to look at evaluation measurements more closely. First, we check to see if the model has any individual p-values above 0.05. If it does, we do not consider it unless they are marginal (i.e. close to 0.05). Next, we look at number of observations. If the model violates the 10:1 data point-to-variable ratio, we do not consider it unless it is marginal (i.e. close to 10:1). Third, we look at average p-values. Finally, we look at adjusted R^2.

Deciding whether a variable will be added to the two-variable model is a combination of the last two evaluation measures. In most cases, the models with the lowest average p-values also yield the highest adjusted R^2, and are therefore easy to pick. In the rare instance where this is not the case, we make a judgment call which results in one of two outcomes: picking one model over another, or carrying both models forward to the next round. We repeat this process until the addition of variables no longer yields favorable results. In this case, we discover no benefit to adding a third variable to the model.

Given the fact that we only have a two-variable model, we run the 73 excluded variables with it and discover that two of them significantly improve the model. First, variable 50 is added, then variable 57. Adding further variables beyond this point offers no improvement. Table 4.7 summarizes the evaluation measurements for the best model of each size (from two to four variables). Unlike our logistic regression models, data point-to-variable ratio is now a bigger factor, as model 4 actually breaches 10:1 threshold, but is still acceptable according to Neter. Thus, we choose model 4 as our final full model because of its high R^2 and very low p-values. Attachment 14 provides the full JMP® analysis of model 4.

Table 4.7 - Best 2-4 Variable OLS Regression Models (Engineering)

Best Models Using 2-4 Variables				
Model#	**# Variables**	**Ave P-Value**	**R^2**	**DP:Var Ratio**
2	2	0.0094	0.2447	20:1
3	3	0.0188	0.3226	13.3:1
4	4	0.0067	0.4717	9.5:1

Now that we have chosen our final full model, we perform the manipulations mentioned in Chapter Three to see if we can make a better, reduced model. This yields us one promising observation: We create variable 83 by dividing variable 50 (funding years of procurement completed) by variable 57 (maturity of EMD%) to make one variable instead of two, which improves our model. This manipulation then allows us to add a previously excluded variable (76), which significantly improves our model. This manipulation achieves this by increasing the data point-to-variable ratio, thus creating "room" for another variable in the model. Table 4.8 compares the final full model to this reduced candidate.

Table 4.8 - Full/Reduced Model Comparison (Engineering)

Full/Reduced Model Comparison				
Model#	**# Variables**	**Ave P-Value**	**R^2**	**DP:Var Ratio**
4	4	0.0067	0.4717	9.5:1
6	4	0.0048	0.5484	8.8:1

Note that the data point-to-variable ratio has dropped to about 9:1. Given the drop in p-values and significant increase in R^2, we deem this ratio to be acceptable and select the reduced candidate to be our final reduced model. To see all six models and their variables, refer to Attachments 3 and 5. Attachment 15 provides the full JMP® analysis on model 6. The predictor variables used in model 6 are:

Quantity Currently Estimated for R&D
Unit Cost
Combined Variable 83 (Fund Years of Proc Completed/Maturity of EMD%)
Program have a MSI?

Next, perform the assumption tests, starting with normality of the residuals. Inspection of Figure 4.5 indicates that we may have a problem with this assumption. Graphically, the distribution looks somewhat normal, but we fail the Shapiro-Wilkes test with a p-value of less than 0.05. We conclude that this is an acceptable outcome for this assumption test, and continue to the assumption of independence.

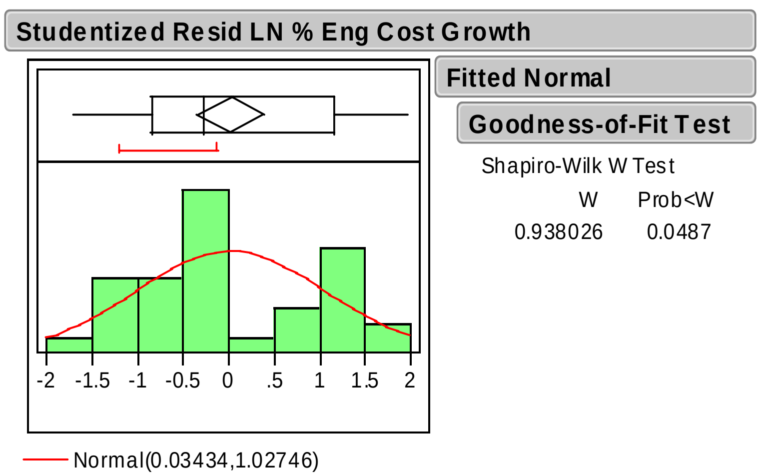


Figure 4.5 - Goodness of Fit Test for Normality (OLS Engineering)

Next, we look at the assumption of independence. Two characteristics of the database lead us to conclude that we pass this assumption. First, the database is made up 135 distinct programs. No programs are repeated. Second, our data was selected randomly. We simply used the latest SAR for each program, and used only programs that had a SAR between 1990 and 2002.

Our final assumption test is for homoscedasticity (constant variance). With a p-value of 0.86, we pass the Breusch-Pagan test for homoscedasticity of the residuals, since the p-value is greater than 0.05.

After completing the assumption tests, we perform a Cook's Distance analysis to check for overly influential data points. The first plot in Figure 4.6 indicates that data point 19 is well over the threshold of 0.5, so we remove it. The next plot is the Cook's Distance result with 19 removed. Now it appears that data point 31 is in the marginal zone (between 0.25 and 0.50) so we remove it and run the model again.

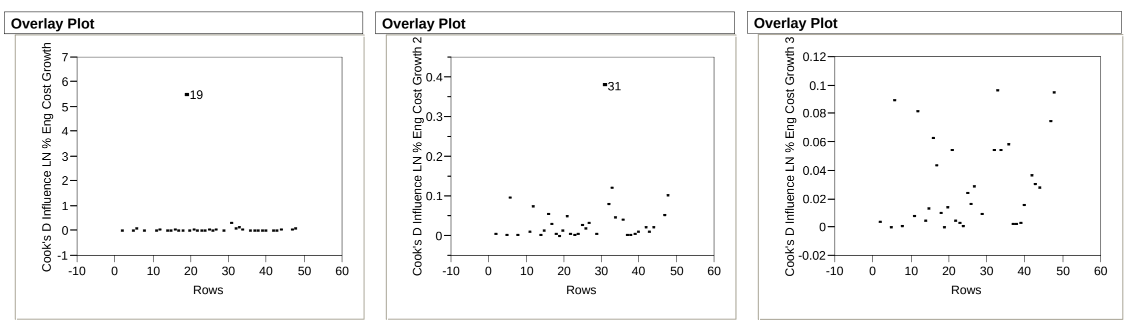


Figure 4.6 - Cook's Distance Plots (OLS Engineering)

The third plot has both 19 and 31 removed, and shows no potential influential data points. The removal of the marginal data point, however, causes the individual p-value of variable 76 jump well over our threshold of 0.05. Given that data point 31 was marginal, and the model performs better with it included, we conclude that the marginal Cook's Distance score is acceptable and the parameters of our model will be generated with only data point 19 excluded. Attachment 16 provides the full JMP® analysis of model 4 with data point 19 excluded. The final step before validation is to check the VIF scores. Figure 4.7 shows that no VIF scores exceed 10, so we proceed to validation.

Parameter Estimates					
Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	-2.29296	0.483138	-4.75	<.0001	.
5 Qty currently estimated for R&D	-0.026373	0.005934	-4.44	0.0001	1.0573576
3 Unit Cost	-0.005825	0.002515	-2.32	0.0278	1.0507865
83 (50/57)	0.137765	0.039515	3.49	0.0016	1.1150278
76 Program have a MS I?	-1.138272	0.483723	-2.35	0.0256	1.1150715

Figure 4.7 - VIF Scores (OLS Engineering)

The final step for our reduced model is to validate it using the 20% of the database that we set aside prior to model building. We do this by putting the entire database back together, running our reduced model, and observing how many times the actual cost growth falls below our predicted 90th percentile.

Of the 12 data points we use in the validation, nine are usable (75%). Out of all 61 data points in the database, 44 are usable (72.1%). From this, we conclude that our model is fairly universal. Of the nine observations we use in the validation, our model yields eight accurate predictions (88.9%). Out of all 44 observations, our model yields 40 accurate predictions (90.1%). From this, we conclude that our model is highly effective at predicting cost growth below the 90th percentile.

In conclusion, our OLS regression model to predict cost growth does have highly predictive capability as it produces 90.1% accurate predictions up to the 90th percentile, and is also fairly universal, as 72.1% of the data points in our database are usable.

Schedule Cost Category

The second model we build uses the natural log of the Schedule Cost Growth % response variable. In round one, our regressions yield a total of ten foundation variables. Round two eliminates one foundation variable (it does not perform well with any other variable) and leaves us with nine two-variable models. The results to this point also allow us to exclude 59 of the 81 predictor variables, as they do not achieve average p-values below 0.05 by themselves, or with any other variable.

We repeat the process from round three on, leaving us with four models ranging from two to five variables in size. Table 4.9 summarizes the evaluation measurements for these models. Although model 33 has a data point-to-variable ratio of about 9:1 (less than 10:1), we select it as our final full model because of its high R^2 and extremely low p-values. Attachment 17 provides the full JMP® analysis of model 33.

Table 4.9 - Best 2-4 Variable OLS Regression Models (Schedule)

Best Models Using 2-5 Variables				
Model#	**# Variables**	**Ave P-Value**	**R^2**	**DP:Var Ratio**
9	2	0.0002	0.2447	24.5:1
11	3	0.0008	0.3216	15.3:1
21	4	0.0036	0.5300	11.5:1
33	5	0.0020	0.5811	9.2:1

Manipulation of model 33 yields us one promising observation: We discretize variable 50 by making any values greater than 9.5 true, which improves our model. Table 4.10 compares the final full model to this reduced candidate.

Table 4.10 - Full/Reduced Model Comparison (Schedule)

Full/Reduced Model Comparison				
Model#	**# Variables**	**Ave P-Value**	**R^2**	**DP:Var Ratio**
33	5	0.0020	0.5811	9.2:1
37	5	0.0019	0.6187	9.2:1

Given the drop in p-values and increase in R^2, we would have selected model 37 to be our final reduced model. But as we move on to the Cook's Distance analysis, we are forced to remove a data point which causes a linear redundancy in the intercept and one of the independent variables. This requires forcing the regression line through the origin to overcome this difficulty. We are unable to use model 37, as a result, and model 33 becomes our final reduced model. To see all 37 models and their variables, refer to Attachments 4 and 5. Attachment 18 provides the full JMP® analysis on model 33. The predictor variables used in model 6 are:

Procurement Started Based on Funding Years?
Lockheed Martin?
Space (Rand Definition)
Funding Years of Procurement Completed
Navy Involvement?

Next, we perform the assumption tests, starting with normality of the residuals. Inspection of Figure 4.8 indicates that we pass this assumption, both graphically, and with the Shapiro Wilke's p-value of 0.62.

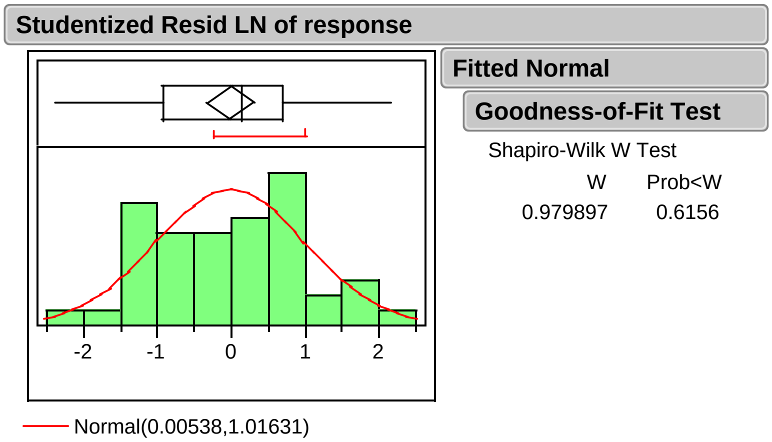


Figure 4.8 - Goodness of Fit Test for Normality (OLS Schedule)

Next, we look at the assumption of independence. Two characteristics of the database lead us to conclude that we pass this assumption. First, the database is made up 135 distinct programs. No programs are repeated. Second, our data was selected randomly. We simply used the latest SAR for each program, and used only programs that had a SAR between 1990 and 2002.

Our final assumption test is for homoscedasticity (constant variance). With a p-value of 0.28, we pass the Breusch-Pagan test for homoscedasticity of the residuals, since the p-value is greater than 0.05.

After completing the assumption tests, we perform a Cook's Distance analysis to check for overly influential data points. The first plot in figure 4.9 indicates that data point 46 is well over the threshold of 0.5, so we remove it. The second plot is the Cook's Distance results with 46 removed, which indicates that we no longer have any points above 0.5, nor any marginal points above 0.25.

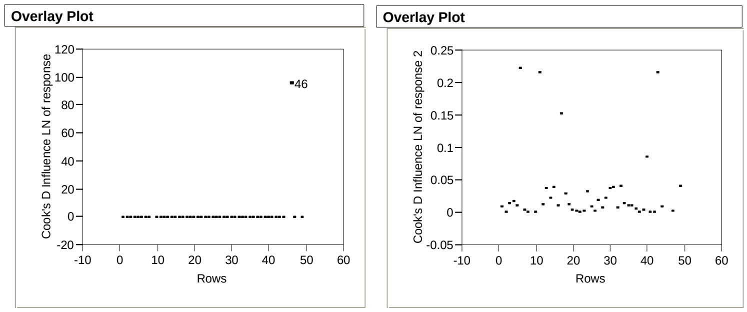


Figure 4.9 - Cook's Distance Plots (OLS Schedule)

The removal of the overly influential data points, however, causes the average p-values to increase, and the adjusted R^2 to decrease, but no to the extent that the model should be thrown out. Attachment 19 provides the full JMP® analysis of model 33 with data point 46 excluded. The final step before validation is to check the VIF scores. Figure 4.10 shows that no VIF scores exceed 10, so we proceed to validation.

Parameter Estimates

Term		Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	Biased	-4.373885	0.432744	-10.11	<.0001	.
62 Proc Started based on Funding Yrs?	Zeroed	0	0	.	.	0
38 Lockheed-Martin		-1.686256	0.468593	-3.60	0.0009	1.2542862
18 Space (RAND)		2.944789	0.981521	3.00	0.0046	1.2654116
50 Funding Yrs of Proc Completed		0.1187131	0.035013	3.39	0.0016	1.0934752
35 N involvement		1.223305	0.382737	3.20	0.0027	1.1052538

Figure 4.10 - VIF Scores (OLS Schedule)

Finally, we validate our model to see how many times the actual cost growth falls below our predicted 90th percentile. Of the 12 data points we use in the validation, 12 are usable (100%). Out of all 61 data points in the database, 58 are usable (95.1%). From this we conclude that our model is highly universal. Of the 12 observations we use in the validation, our model yields 10 accurate predictions (83.3%). Out of all 58 observations, our model yields 52 accurate predictions (89.7%). From this, we conclude that our model is highly effective at predicting cost growth below the 90^{th} percentile.

In conclusion, our OLS regression model to predict cost growth does have highly predictive capability as it produces 89.7% accurate predictions up to the 90^{th} percentile, and is also highly universal, as 72.1% of the data points in our database were able to yield predictions.

V. Conclusions

Justification for Research

Defense spending has experienced increased scrutiny as the high levels of spending of the Reagan Administration gave way to record-setting reductions during the Clinton Administration. The threats that defense spending are designed to mitigate, however, have not reduced. This results in a change of business practice in the defense acquisition community, which faces the challenge of retaining its credibility in the eyes of Congress and the American taxpayer. To this point, the defense acquisition community has failed to retain its credibility, as despite numerous legislative reforms, cost growth on major defense acquisition programs has increased 9.9% since 1990 (Suddarth, 2002:7).

DoD leadership sees improvements in cost estimates as a logical way to decrease cost growth. This philosophy is not new, as DoD directives have stated for years that the cost of major acquisition programs is to be considered an independent variable and that the goal is to estimate cost with 60-70% confidence. Dr. Sambur and Lieutenant General Keys increased this confidence level to 90% while addressing the House and Armed Services Committee in 2003.

The purpose of this research is to build regression models that can be used to predict cost growth with a 90% level of confidence so that major defense acquisition programs can be properly funded up front. Such a business practice would decrease overall cost growth in the defense acquisition community, and restore credibility to its executives.

Limitations

This research has three important limitations. First, it only addresses the Engineering and Schedule cost categories. Cost growth has two other categories that program managers have influence over: Estimating and Support. Therefore, predictions made using our models will not reflect cost growth or cost savings in the other two categories.

Second, our models are based on the relationships between variables in the EMD phase, and only make predictions for cost growth during that phase. It is not reasonable to assume that these models can be used to predict cost growth outside of this phase.

Third, these models only predict cost growth for efforts funded with procurement dollars. It is not reasonable to assume that these models can be applied to efforts funded with other types of appropriated dollars.

Finally, the strength of these models is limited to the unclassified portions of the SAR, and only programs that submitted a SAR from 1990-2002. Relationships between classified variables or variables from programs that did not submit a SAR during this timeframe, are not used in the model-building process.

Review of Methodology

We validate and update the database originally developed by Sipple using the most current SAR database. The data must be unclassified and from mature programs (i.e. at least three years into the EMD phase). Once the database is updated, we review the predictor variables and delete, modify and add as needed.

Next, we review the distributions of the response variables, and conclude that we, like Sipple, Bielecki and Moore, have a mixed-distribution with a discrete mass at zero, which violates the assumption that our responses are reasonably continuous. Therefore, we also elect to use the two-pronged logistic and multiple regression approach. Next, we confirm that the logistic and multiple response variables have enough responses and variety in the responses to build sound models. Finally, we partition the database: 80% for model building, and 20% for validation. Then we proceed to model building.

First, we build our logistic regression models, which predict whether cost growth will occur in the engineering and schedule categories. We build the models one variable at a time by running combinations of variables through successive rounds, carrying forward models with the best performance in the evaluation measures (below).

- Average individual p-values
- R^2(U)
- Data point-to-variable ratio
- Area under the receiver operating curve (ROC)

We continue until the process of adding variables provides no further benefit. Based on the measures above, we select final full models, and perform manipulations in an attempt to increase their predictive ability. Once we have selected final reduced models, we validate them against the 20% database and the 100% database to determine if they are predictive and universal.

Second, we build our OLS regression models, to predict the level of cost growth that will occur in the engineering and schedule categories. We build the models one variable at a time by running combinations of variables through successive rounds, carrying forward models with the best performance in the evaluation measures listed on the following page.

- Average individual p-values
- Adjusted R^2
- Data point-to-variable ratio

We continue this until the process of adding variables produces no further benefit. Based on the measures above, we select final full models, and perform manipulations in an attempt to increase their predictive ability. Once we have selected final reduced models, we look for overly influential data points and multicollinearity between the predictor variables using the Cook's Distance test and Variance Inflation Factor analysis.

Once we have models that make it to this point, we test to ensure they pass the assumptions of normality of the residuals, independence, and homoscedasticity of the residuals. Once these tests are complete, we validate the models against the 20% database and the 100% database to determine if they are predictive and universal.

Restatement of Results

Logistic Regression

We select model 49 as our final reduced model for the Engineering cost category. We restate evaluation measurements and validation results in Table 5.1. We conclude that this model does have predictive capability as it does accurately predict cost growth for the programs in our 100% database 74.6% of the time, despite a poor validation score. We attribute the poor validation score, at least in part, to a validation set that does not accurately represent the population. We also conclude that it is highly universal, as 93.3% of all the programs studied are usable. Model 49 contains the following predictor variables.

Length of R&D in Funding Years
Classification Secret?
LRIP Planned?
Lockheed Martin?
Discretized Variable 54 (Proc Funding Yr Maturity % >.45?)
Risk Mitigation?

Table 5.1 - Overall Logistic Regression Results (Engineering)

Evaluation Measures for Final Reduced Model				
Model#	# Variables	Ave P-Value	R^2(U)	ROC
49	6	0.0219	0.3264	0.8622
Validation for Final Reduced Model				
Model#	% Observations Validation Set	% Accurate Validation Set	% Observations Full Data Set	% Accurate Full Data Set
49	96.3%	50.0%	93.3%	74.6%

Model 36 is our best logistic regression model for the Schedule cost category, and its evaluation measures and validation results are restated in Table 5.2. We conclude that this model does have predictive capability as it accurately predicts cost growth for the programs in our 100% database 78.1% of the time, despite a poor validation score. We attribute at the poor validation, at least in part to the fact that 20% database does not accurately represent the population. We also conclude that this model is highly universal, as 94.8% of all the programs studied were usable. Model 36 contains the following predictor variables.

Discretized Variable 54 (Proc Funding Yr Maturity % >0.43?)
68 Risk Mitigation?
Marine Core Involvement?
LRIP Planned?
Interaction Term (MC Involvement? and LRIP Planned?)

Table 5.2 - Overall Logistic Regression Results (Schedule)

Evaluation Measures for Final Reduced Model				
Model#	**# Variables**	**Ave P-Value**	**R^2(U)**	**ROC**
36	4	0.0172	0.4392	0.8913
Validation for Final Reduced Model				
Model#	**% Observations Validation Set**	**% Accurate Validation Set**	**% Observations Full Data Set**	**% Accurate Full Data Set**
36	96.3%	50.0%	94.8%	78.1%

<u>*OLS Regression*</u>

We select model 6 as our final reduced model for the Engineering cost category. We restate evaluation measurements and validation results in Table 5.3. We conclude that this model does have predictive capability as it does accurately predict cost growth for the programs in our 100% database 90.1% of the time, and validates at 88.9%. We also conclude that it is fairly universal, as 72.1% of all the programs studied are usable, as well as 75% of the validation set. Model 6 contains the following predictor variables.

Quantity Currently Estimated for R&D
Unit Cost
Combined Variable 83 (Fund Years of Proc Completed/Maturity of EMD%)
Program have a MSI?

Table 5.3 - Overall OLS Regression Results (Engineering)

Evaluation Measures for Final Reduced Model				
Model#	**# Variables**	**Ave P-Value**	**R^2**	**DP:Var Ratio**
6	4	0.0048	0.5484	8.8:1
Validation for Final Reduced Model				
Model#	**% Observations Validation Set**	**% Accurate Validation Set**	**% Observations Full Data Set**	**% Accurate Full Data Set**
6	75.0%	88.9%	72.1%	90.1%

Model 33 is our best final reduced model for the Schedule cost category. We restate evaluation measurements and validation results in table 5.4. We conclude that this model is highly predictive as it accurately predicts cost growth for the programs in our 100% database 89.7% of the time, and validates at 83.3%. We also conclude that it is highly universal, as 95.1% of all the programs studied are usable, as well as 100% of the validation set. Model 33 contains the following predictor variables.

Procurement Started Based on Funding Years?
Lockheed Martin?
Space (Rand Definition)
Funding Years of Procurement Completed
Navy Involvement?

Table 5.4 - Overall OLS Regression Results (Schedule)

Evaluation Measures for Final Reduced Model				
Model#	**# Variables**	**Ave P-Value**	**R^2**	**DP:Var Ratio**
33	5	0.0020	0.5811	9.2:1
Validation for Final Reduced Model				
Model#	**% Observations Validation Set**	**% Accurate Validation Set**	**% Observations Full Data Set**	**% Accurate Full Data Set**
33	100.0%	83.3%	95.1%	89.7%

<u>*Comparison to Moore's Models*</u>

Being that Moore's thesis is the only work involving procurement dollars, it may be useful to compare his models to ours. Table 5.5 summarizes the evaluation measures and validations of the three sets of procurement-based models (two sets of ours and one set of Moore's). We make two important points for standardization purposes: First, we calculate data point-to-variable ratios using the number of variables used in a model, excluding interaction terms. Second, average p-values take into consideration interactions terms, but they do not include the variables that make up those interaction terms. Third, validations on OLS models can be compared, even though we use a 90%

prediction bound while Moore's thesis uses 80%. This is because Moore calculated his 80% upper bound as if we were using a two-tailed interval, which we are not. Since we consider any prediction below the upper bound to be accurate, Moore is actually validating at 90% because he includes the lower 10% tail as an accurate prediction. Therefore, we are both validating at 90%, and can compare results.

Table 5.5 - Comparison With Moore Thesis

Evaluation Measures for Logisitic Regression					
Response Variable	**# Variables**	**Ave P-Value**	**R^2(U)**	**ROC**	**DP/Var Ratio**
Engineering Cost Growth?	6	0.0219	0.3264	0.8622	16.7:1
Schedule Cost Growth?	4	0.0053	0.4392	0.8913	25.5:1
Overall Cost Growth?	3	0.0578	0.8307	0.9930	11.7:1
Validations for Logistic Regression Models					
Response Variable	**% Observations Validation Set**	**% Accurate Validation Set**	**% Observations Full Data Set**	**% Accurate Full Data Set**	
Engineering Cost Growth?	96.3%	50.0%	93.3%	74.6%	
Schedule Cost Growth?	96.3%	50.0%	94.8%	78.1%	
Overall Cost Growth?	16.0%	100.0%	32.0%	94.9%	
Evaluation Measures for OLS Regression					
Response Variable	**# Variables**	**Ave P-Value**	**Adj R^2**	**DP/Var Ratio**	
Engineering Cost Growth%	4	0.0048	0.5484	8.8:1	
Schedule Cost Growth%	5	0.0020	0.5811	9.2:1	
Overall Cost Growth%	3	0.0096	0.5946	7.3:1	
Validations for OLS Models					
Response Variable	**% Observations Validation Set**	**% Accurate Validation Set**	**% Observations Full Data Set**	**% Accurate Full Data Set**	
Engineering Cost Growth%	75.0%	88.9%	72.1%	90.1%	
Schedule Cost Growth%	100.0%	83.3%	95.1%	89.7%	
Overall Cost Growth%	23.5%	100.0%	34.7%	100.0%	

Moore's logistic regression model outperforms our models on R^2(U) and ROC scores, but has higher p-values. His data point-to-variable ratio is noticeably lower, but still at an acceptable level. His validation accuracy is much higher than both our models, but use a much smaller number of data points. Overall, I would say that despite Moore's impressive R^2(U), ROC and validation accuracy scores, his logistic regression model may have limited real-world utility. The main reason for this is that the number of data points used to generate his model is small compared to the number of data points available.

Moore's OLS regression model slightly outperforms our models on adjusted R^2 and ROC but has higher p-values. His data point-to-variable ratio is lower, but still at an acceptable level (according to Neter). His validation accuracy is much higher than both our models, but use a much smaller number of data points. In keeping with our comments in the previous paragraph, we would say that despite Moore's higher adjusted R^2 and perfect validation accuracy scores, his model may have limited real-world utility due to the limited use of the available data points.

Recommendations

The results from this study further validate the potential for logistic and OLS regression in defense acquisition community. Specifically, this two-pronged regression approach is effective at predicting if cost growth will occur, and very effective at providing a 90% prediction level for budgeting purposes. This could greatly reduce cost overruns in the Department of Defense, restoring credibility to the executives of the defense acquisition community.

In employing this recommendation, it is important to note the importance of the two-pronged approach, as neither this study nor its predecessors encounter a percent cost growth response variable with a reasonably continuous distribution. This translates into constant variance problems for those who attempt to use straight OLS regression to predict cost growth in their programs.

Possible Follow-on Theses

- Allow data to build under the new A B C Acquisitions Milestone Phases, then expand the database and perform the same methodology

- Explore a way to convert the old I II III Milestone phased data into the new A B C phased data

- Identify programs that did not have significant overruns and evaluate their risk estimating methodology to see if there is a best methodology (Sipple, 2002:121)

- Build regression models to predict zero cost growth or cost savings, to determine the characteristics of highly successful programs

- Create a program utilizing the CERs developed from this and other analyses (Sipple, 2002:121)

- Explore the applicability of our results to the Monte Carlo simulation technique of risk analysis (Sipple, 2002:121)

Attachment 1

Best Logistic Regression Models (Engineering Cost Category)

Model #	*Variables*	*# Variables*	*Cum Ind P-Values*	*Ave P-Values*	*R2(U)*	*ROC*
1	51,68	2	0.0198	0.0099	0.1053	0.7236
2	50,77	2	0.0034	0.0017	0.1517	0.7474
3	80,77	2	0.0076	0.0038	0.1296	0.7348
4	81,12	2	0.0140	0.0070	0.1141	0.7162
5	47,77	2	0.0152	0.0076	0.0985	0.6907
6	54,77	2	0.0048	0.0024	0.1113	0.6779
7	79,49	2	0.0142	0.0071	0.1057	0.7222
8	52,12	2	0.0164	0.0082	0.1015	0.7097
9	62,12	2	0.0187	0.0094	0.1019	0.6790
10	48,54	2	0.0367	0.0184	0.0832	0.6909
11	73,71	2	0.0119	0.0060	0.0864	0.6722
12	7,68	2	0.0368	0.0184	0.0987	0.6778
13	51,68,54	3	0.0490	0.0163	0.1434	0.7475
14	50,77,68	3	0.0315	0.0105	0.2019	0.7758
15	80,77,65	3	0.0221	0.0074	0.1693	0.7638
16	81,12,65	3	0.0276	0.0092	0.1535	0.7509
17	47,77,65	3	0.0195	0.0065	0.1422	0.7312
18	54,77,81	3	0.0198	0.0066	0.1581	0.7478
19	79,49,12	3	0.0465	0.0155	0.1454	0.7514
20	52,12,65	3	0.0253	0.0084	0.1423	0.7431
21	62,12,73	3	0.0508	0.0169	0.1336	0.7281
22	48,54,77	3	0.0359	0.0120	0.1446	0.7338
23	73,71,50	3	0.0097	0.0032	0.1772	0.7769
24	7,68,65	3	0.0535	0.0178	0.1339	0.7403
25	51,68,54,77	4	0.0363	0.0091	0.2132	0.7968
26	50,77,68,65	4	0.0661	0.0165	0.2327	0.8026
27	80,77,65,68	4	0.0523	0.0131	0.2359	0.8094
28	81,12,65,18	4	0.0404	0.0101	0.1944	0.7829
29	47,77,65,68	4	0.0715	0.0179	0.2042	0.7838
30	54,77,81,68	4	0.0357	0.0089	0.2226	0.8034
31	79,49,12,77	4	0.0498	0.0125	0.1720	0.7600
32	52,12,65,77	4	0.0318	0.0080	0.1828	0.7691
33	62,12,73,71	4	0.0447	0.0112	0.1846	0.7723
34	48,54,77,65	4	0.0413	0.0103	0.1802	0.7672
35	73,71,50,77	4	0.0405	0.0101	0.2281	0.8044
36	7,68,65,77	4	0.0508	0.0127	0.1760	0.7836
37	51,68,54,77,65	5	0.0771	0.0154	0.2435	0.8123
38	54,77,81,68,65	5	0.0556	0.0111	0.2620	0.8292
39	79,49,12,77,65	5	0.0732	0.0146	0.2086	0.7898
40	52,12,65,77,38	5	0.0758	0.0152	0.2393	0.8103
41	62,12,73,71,38	5	0.0825	0.0165	0.1985	0.7814
42	48,54,77,65,68	5	0.0566	0.0113	0.2464	0.8131
43	73,71,50,77,68	5	0.0873	0.0175	0.2505	0.8172
44	7,68,65,77,50	5	0.1228	0.0246	0.2585	0.8278
45	54,77,81,68,65,38	6	0.1455	0.0243	0.2760	0.8348
46	52,12,65,77,38,54	6	0.1506	0.0251	0.2705	0.8259
47	54,77,81,68,65,38,12	7	0.3067	0.0438	0.2930	0.8394
48	52,12,65,77,38,54,68	7	0.2420	0.0346	0.2992	0.8473
49	52,65,77,38,82,68	6	0.1312	0.0219	0.3264	0.8622

Color Key:

Final Full Model

Final Reduced Model

Attachment 2

Best Logistic Regression Models (Schedule Cost Category)

Model #	*Variables*	*# Variables*	*Cum Ind P-Values*	*Ave P-Values*	*R2(U)*	*ROC*
1	54,68	2	0.0012	0.0006	0.2696	0.8356
2	54,74	2	0.0097	0.0049	0.1693	0.7632
3	50,68	2	0.0040	0.0020	0.2063	0.8189
4	68,80	2	0.0086	0.0043	0.1660	0.7663
5	68,62	2	0.0116	0.0058	0.2057	0.7325
6	80,46	2	0.0339	0.0170	0.1138	0.7450
7	62,46	2	0.0510	0.0255	0.1525	0.7009
8	47,46	2	0.0369	0.0185	0.1023	0.7165
9	64,68	2	0.0622	0.0311	0.1339	0.7063
10	47,68	2	0.0225	0.0113	0.1445	0.7448
11	46,50	2	0.0300	0.0150	0.1359	0.7816
12	46,54	2	0.0223	0.0112	0.1848	0.7708
13	49,46	2	0.0605	0.0303	0.0812	0.6992
14	55,68	2	0.0199	0.0100	0.1445	0.7556
15	57,68	2	0.0534	0.0267	0.1509	0.7374
16	64,68,36	3	0.0296	0.0099	0.3062	0.8478
17	54,74,68	3	0.0344	0.0115	0.3041	0.8521
18	50,68,62	3	0.0743	0.0248	0.2481	0.8313
19	68,80,49	3	0.0179	0.0060	0.2171	0.8195
20	68,62,55	3	0.0319	0.0106	0.2467	0.8267
21	62,46,55	3	0.0612	0.0204	0.1922	0.7829
22	47,46,68	3	0.0942	0.0314	0.1885	0.7784
23	64,68,11	3	0.0750	0.0250	0.1747	0.7420
24	47,68,79	3	0.0467	0.0156	0.1864	0.7640
25	46,50,74	3	0.0504	0.0168	0.1752	0.8028
26	46,54,74	3	0.0277	0.0092	0.2433	0.8137
27	55,68,79	3	0.0059	0.0020	0.2314	0.8059
28	54,68,36,77	4	0.0651	0.0163	0.3375	0.8667
29	68,62,55,36	4	0.0625	0.0156	0.2785	0.8439
30	68,62,55,11	4	0.0789	0.0197	0.2818	0.8414
31	62,46,55,36	4	0.0884	0.0221	0.2298	0.8070
32	62,46,55,74	4	0.0681	0.0170	0.2299	0.8011
33	46,50,74,62	4	0.1118	0.0280	0.2232	0.8096
34	55,68,79,46	4	0.0738	0.0185	0.2627	0.8266
35	82,68,36,77	4	0.0614	0.0154	0.3595	0.8699
36	82,68,36,77,(36,77)*	5	0.0860	0.0172	0.4392	0.8913

Color Key:

Final Full Model
Final Reduced Model

* Parentheses indicate interaction term

Attachment 3

Best OLS Models (Engineering Cost Category)

Model #	*Variables*	*# Variables*	*Cum Ind P-Values*	*Ave P-Values*	*Adj R2*
1	7,16	2	0.0681	0.0341	0.1461
2	5,3	2	0.0188	0.0094	0.2447
3	5,3,50	3	0.0354	0.0118	0.3226
4	5,3,50,57	4	0.0268	0.0067	0.4717
5	5,3,50,76	4	0.0399	0.0100	0.4491
6	5,3,83,76	4	0.0190	0.0048	0.5484

Color Key:

Final Full Model
Final Reduced Model

Attachment 4

Best OLS Models (Schedule Cost Category)

Model #	*Variables*	*# Variables*	*Cum Ind P-Values*	*Ave P-Values*	*Adj R2*
1	50,43	2	0.0094	0.0047	0.3105
2	54,23	2	0.0093	0.0047	0.2207
3	43,38	2	0.0096	0.0048	0.2321
4	62,38	2	0.0080	0.0040	0.2793
5	19,50	2	0.0172	0.0086	0.1943
6	38,62	2	0.0080	0.0040	0.2791
7	55,23	2	0.0241	0.0121	0.1625
8	57,38	2	0.0710	0.0355	0.1654
9	1,50	2	0.0003	0.0002	0.3216
10	1,80	2	0.0007	0.0004	0.2844
11	50,43,38	3	0.0024	0.0008	0.4496
12	50,43,1	3	0.0098	0.0033	0.4486
13	54,23,38	3	0.0580	0.0193	0.3249
14	43,38,80	3	0.0175	0.0058	0.4219
15	62,38,18	3	0.0032	0.0011	0.4061
16	62,38,50	3	0.0080	0.0027	0.4197
17	19,50,35	3	0.0037	0.0012	0.3368
18	80,1,8	3	0.0199	0.0066	0.3927
19	55,23,43	3	0.0721	0.0240	0.2661
20	1,50,8	3	0.0098	0.0033	0.4486
21	50,43,38,1	4	0.0145	0.0036	0.5300
22	50,43,1,35	4	0.0150	0.0038	0.5283
23	50,23,38,43	4	0.0517	0.0129	0.4892
24	43,38,80,18	4	0.0195	0.0049	0.4574
25	62,38,18,50	4	0.0245	0.0061	0.4870
26	62,38,50,1	4	0.0212	0.0053	0.5131
27	19,50,35,42	4	0.4349	0.1087	0.4348
28	80,1,8,35	4	0.0600	0.0150	0.4377
29	1,50,8,35	4	0.0291	0.0073	0.4989
30	50,43,38,1,35	5	0.0711	0.0142	0.5696
31	50,23,38,43,18	5	0.0484	0.0097	0.5490
32	43,38,80,18,1	5	0.0593	0.0119	0.5157
33	62,38,18,50,35	5	0.0099	0.0020	0.5811
34	62,38,50,1,18	5	0.0554	0.0111	0.5517
35	19,50,35,42,1	5	0.0141	0.0028	0.5696
36	1,50,8,35,56	5	0.0427	0.0085	0.5512
37	62,38,18,35,82	5	0.0096	0.0019	0.6187

Color Key:

Final Full Model
Final Reduced Model

Attachment 5

List of Original Variables (by number)

1 Total Cost CY $M 2002
2 Total Quantity
3 Unit Cost
4 Qty planned for R&D
5 Qty currently estimated for R&D
6 ACAT
7 ACAT 1?
8 Air
9 Land
10 Space
11 Sea
12 Electronic
13 Helo
14 Missile
15 Aircraft
16 Munition
17 Land Vehicle
18 Space (RAND)
19 Ship
20 Other
21 # of Svs
22 Svs>1
23 Svs >2
24 Svs>3
25 Service = Navy only
26 Service = Joint
27 Service = Army only
28 Service = Marines only
29 Service = AF only
30 Lead Svc = Army
31 Lead Svc = Navy
32 Lead Svc = DoD
33 Lead Svc = AF
34 AF involvement
35 N involvement
36 MC involvement
37 AR involvement
38 Lockheed-Martin
39 Northrop Grumman
40 Boeing
41 Raytheon
42 Litton
43 General Dynamics
44 No Major Defense Contractor
45 More than 1 Major Defense Contractor
46 Fixed-Price EMD Contract?
47 Maturity (Funding Yrs complete)
48 Funding YR Total Program Length
49 Funding Yrs of R&D Completed
50 Funding Yrs of Proc Completed
51 Length of Proc in Funding Yrs
52 Length of R&D in Funding Yrs
53 R&D Funding Yr Maturity %
54 Proc Funding Yr Maturity %
55 Total Funding Yr Maturity %
56 Actual Length of EMD
57 Maturity of EMD %
58 Time from MSII to IOC (in months)
59 Maturity of EMD at IOC%
60 LRIP Qty Planned
61 LRIP Qty Current Estimate
62 Proc Started based on Funding Yrs?
63 Proc Funding before MS III?
64 # Product variants in this SAR
65 Class - S
66 Class - C
67 Class - U
68 Risk Mitigation?
69 Versions Previous to SAR
70 Modification?
71 Prototype?
72 Dem/Val Prototype?
73 EMD Prototype?
74 PE ?
75 Significant pre-EMD activity?
76 Program have a MS I?
77 LRIP Planned?
78 % R&D of Total Program (years)
79 % Proc of Total Program (years)
80 Fund Years of R&D + Prod Complete
81 Length of R&D + Prod Funding Years

Attachment 6

JMP 5.1® Statistical Analysis:
Final Full Logistic Regression Model (Engineering Cost Category)

Nominal Logistic Fit for Engineering Cost Growth? Procurement

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	20.591359	7	41.18272	<.0001
Full	48.222523			
Reduced	68.813881			

RSquare (U)	0.2992
Observations (or Sum Wgts)	100

Converged by Gradient

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	92	48.222523	96.44505
Saturated	99	0.000000	Prob>ChiSq
Fitted	7	48.222523	0.3551

Parameter Estimates

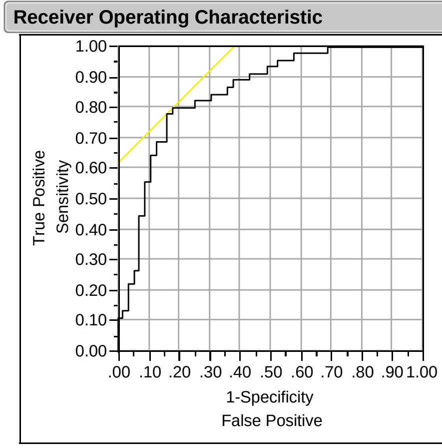
Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	3.55330243	1.166484	9.28	0.0023
52 Length of R&D in Funding Yrs	-0.0937586	0.0344455	7.41	0.0065
12 Electronic	1.03085856	0.5868404	3.09	0.0790
65 Class - S	1.30836431	0.5693017	5.28	0.0216
77 LRIP Planned?	-1.9521659	0.6470295	9.10	0.0026
38 Lockheed-Martin	1.29615088	0.6270468	4.27	0.0387
54 Proc Funding Yr Maturity %	-1.8114624	0.8661215	4.37	0.0365
68 Risk Mitigation?	-1.3463647	0.7075558	3.62	0.0571

For log odds of 0/1

Effect Wald Tests

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
52 Length of R&D in Funding Yrs	1	1	7.40895749	0.0065
12 Electronic	1	1	3.08573186	0.0790
65 Class - S	1	1	5.28168495	0.0216
77 LRIP Planned?	1	1	9.10301612	0.0026
38 Lockheed-Martin	1	1	4.27278663	0.0387
54 Proc Funding Yr Maturity %	1	1	4.37422425	0.0365
68 Risk Mitigation?	1	1	3.62079545	0.0571

Receiver Operating Characteristic



Using Engineering Cost Growth? Procurement='1' to be the positive level
Area Under Curve = 0.84727

Attachment 7

JMP 5.1® Statistical Analysis:
Final Reduced Logistic Regression Model (Engineering Cost Category)

Nominal Logistic Fit for Engineering Cost Growth? Procurement

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	22.461072	6	44.92214	<.0001
Full	46.352809			
Reduced	68.813881			

RSquare (U)	0.3264
Observations (or Sum Wgts)	100

Converged by Gradient

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	81	39.944585	79.88917
Saturated	87	6.408224	Prob>ChiSq
Fitted	6	46.352809	0.5140

Parameter Estimates

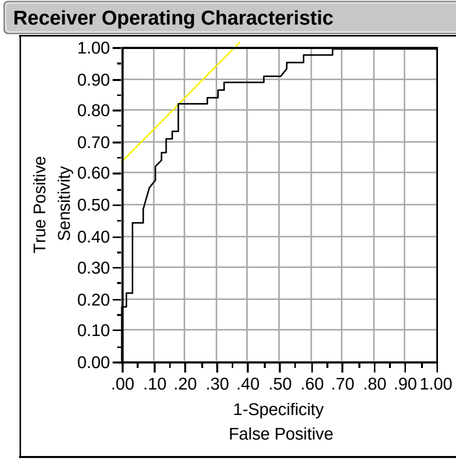
Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	4.58102782	1.154231	15.75	<.0001
52 Length of R&D in Funding Yrs	-0.0863098	0.0365042	5.59	0.0181
65 Class - S	1.16175564	0.5855777	3.94	0.0473
77 LRIP Planned?	-2.6021887	0.7626746	11.64	0.0006
38 Lockheed-Martin	1.32128903	0.6677042	3.92	0.0478
82 Discrete54	-2.4572251	0.7552747	10.58	0.0011
68 Risk Mitigation?	-1.6539073	0.6888368	5.76	0.0163

For log odds of 0/1

Effect Wald Tests

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
52 Length of R&D in Funding Yrs	1	1	5.59028211	0.0181
65 Class - S	1	1	3.93604978	0.0473
77 LRIP Planned?	1	1	11.641233	0.0006
38 Lockheed-Martin	1	1	3.91586242	0.0478
82 Discrete54	1	1	10.584736	0.0011
68 Risk Mitigation?	1	1	5.76487132	0.0163

Receiver Operating Characteristic



Using Engineering Cost Growth? Procurement='1' to be the positive level
Area Under Curve = 0.86222

Attachment 8

JMP 5.1® Statistical Analysis:
Final Full Logistic Regression Model (Schedule Cost Category)

Nominal Logistic Fit for Schedule Cost Growth? Procurement

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	23.698944	4	47.39789	<.0001
Full	46.511084			
Reduced	70.210028			

RSquare (U)	0.3375
Observations (or Sum Wgts)	102

Converged by Gradient

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	58	35.564077	71.12815
Saturated	62	10.947008	Prob>ChiSq
Fitted	4	46.511084	0.1154

Parameter Estimates

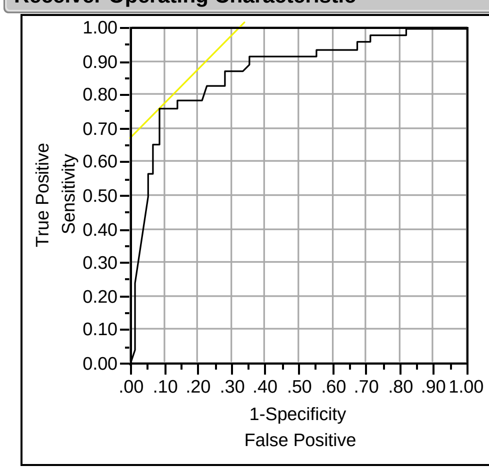
Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	5.65003797	1.2581802	20.17	<.0001
54 Proc Funding Yr Maturity %	-4.2019332	0.912788	21.19	<.0001
68 Risk Mitigation?	-2.9193708	0.9350232	9.75	0.0018
36 MC involvement	-1.4916331	0.6274797	5.65	0.0174
77 LRIP Planned?	-1.1846196	0.5930038	3.99	0.0458

For log odds of 0/1

Effect Wald Tests

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
54 Proc Funding Yr Maturity %	1	1	21.1913413	0.0000
68 Risk Mitigation?	1	1	9.7484086	0.0018
36 MC involvement	1	1	5.65099103	0.0174
77 LRIP Planned?	1	1	3.99064321	0.0458

Receiver Operating Characteristic



Using Schedule Cost Growth? Procurement='1' to be the positive level
Area Under Curve = 0.86665

Attachment 9

JMP 5.1® Statistical Analysis:
Final Reduced Logistic Regression Model (Schedule Cost Category)

Nominal Logistic Fit for Schedule Cost Growth? Procurement

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	30.838620	5	61.67724	<.0001
Full	39.371408			
Reduced	70.210028			

RSquare (U)	0.4392
Observations (or Sum Wgts)	102

Converged by Gradient

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	8	2.505622	5.011245
Saturated	13	36.865785	Prob>ChiSq
Fitted	5	39.371408	0.7564

Parameter Estimates

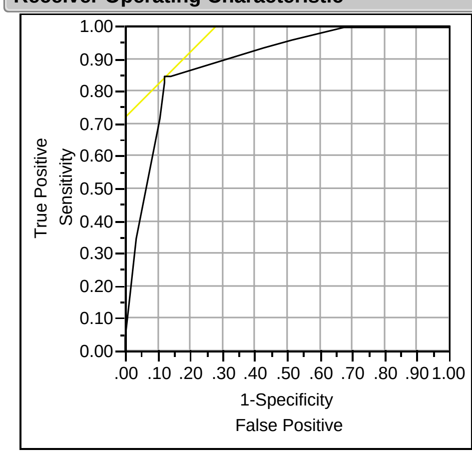
Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	5.33141954	1.1180597	22.74	<.0001
82 Discrete 54	-3.3604978	0.6850695	24.06	<.0001
68 Risk Mitigation?	-3.035857	0.9108291	11.11	0.0009
36 MC involvement	-1.5799626	0.7776993	4.13	0.0422
77 LRIP Planned?	-1.4435912	0.6976474	4.28	0.0385
(36 MC involvement-0.22549)*(77 LRIP Planned?-0.45098)	-4.7406814	1.6596961	8.16	0.0043

For log odds of 0/1

Effect Wald Tests

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
82 Discrete 54	1	1	24.0623446	0.0000
68 Risk Mitigation?	1	1	11.1093553	0.0009
36 MC involvement	1	1	4.12733932	0.0422
77 LRIP Planned?	1	1	4.28170208	0.0385
36 MC involvement*77 LRIP Planned?	1	1	8.15876443	0.0043

Receiver Operating Characteristic

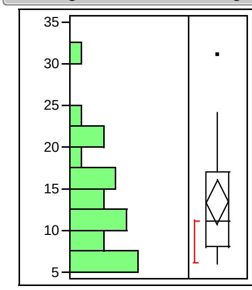


Using Schedule Cost Growth? Procurement='1' to be the positive level
Area Under Curve = 0.89130

Attachment 10

JMP 5.1® Statistical Analysis:
20% Database Variable Distributions (Engineering Cost Category)

52 Length of R&D in Funding Yrs



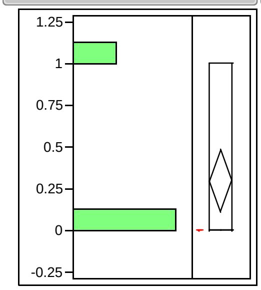
Quantiles

100.0%	maximum	31.000
99.5%		31.000
97.5%		31.000
90.0%		21.600
75.0%	quartile	17.000
50.0%	median	11.000
25.0%	quartile	8.000
10.0%		6.000
2.5%		6.000
0.5%		6.000
0.0%	minimum	6.000

Moments

Mean	13.296296
Std Dev	6.4439287
Std Err Mean	1.2401346
upper 95% Mean	15.84543
lower 95% Mean	10.747163
N	27

65 Class - S



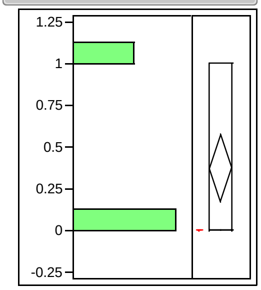
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	0.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.2962963
Std Dev	0.4653216
Std Err Mean	0.0895512
upper 95% Mean	0.4803714
lower 95% Mean	0.1122212
N	27

77 LRIP Planned?



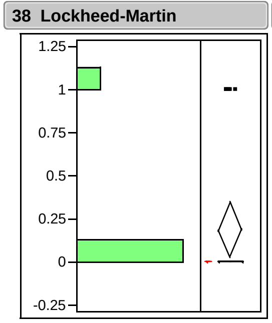
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	0.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.3703704
Std Dev	0.4921029
Std Err Mean	0.0947052
upper 95% Mean	0.5650398
lower 95% Mean	0.175701
N	27

38 Lockheed-Martin



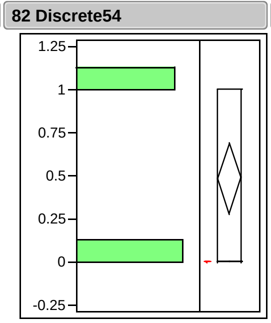
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	0.0000
50.0%	median	0.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.1851852
Std Dev	0.3958474
Std Err Mean	0.0761809
upper 95% Mean	0.3417772
lower 95% Mean	0.0285932
N	27

82 Discrete54



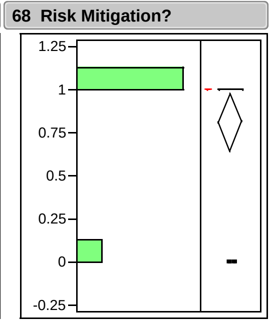
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	0.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.4814815
Std Dev	0.5091751
Std Err Mean	0.0979908
upper 95% Mean	0.6829044
lower 95% Mean	0.2800585
N	27

68 Risk Mitigation?



Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	1.0000
25.0%	quartile	1.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

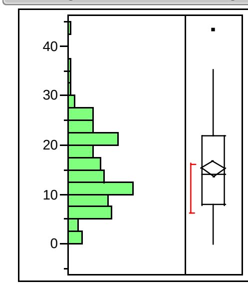
Moments

Mean	0.8076923
Std Dev	0.4019185
Std Err Mean	0.0788227
upper 95% Mean	0.9700307
lower 95% Mean	0.6453539
N	26

Attachment 11

JMP 5.1® Statistical Analysis:
80% Database Variable Distributions (Engineering Cost Category)

52 Length of R&D in Funding Yrs



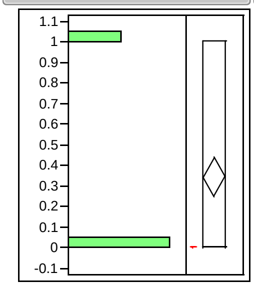
Quantiles

100.0%	maximum	43.000
99.5%		43.000
97.5%		33.550
90.0%		26.000
75.0%	quartile	21.750
50.0%	median	14.000
25.0%	quartile	8.000
10.0%		6.000
2.5%		1.725
0.5%		0.000
0.0%	minimum	0.000

Moments

Mean	15.101852
Std Dev	8.1228187
Std Err Mean	0.7816186
upper 95% Mean	16.65132
lower 95% Mean	13.552384
N	108

65 Class - S



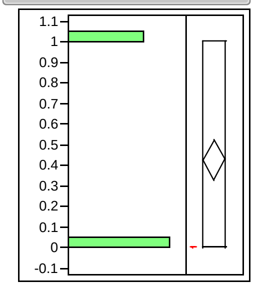
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	0.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.3425926
Std Dev	0.476789
Std Err Mean	0.045879
upper 95% Mean	0.4335425
lower 95% Mean	0.2516427
N	108

77 LRIP Planned?



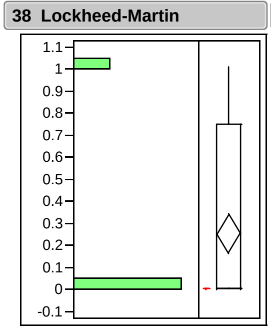
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	0.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.4259259
Std Dev	0.4967879
Std Err Mean	0.0478034
upper 95% Mean	0.5206907
lower 95% Mean	0.3311612
N	108

38 Lockheed-Martin



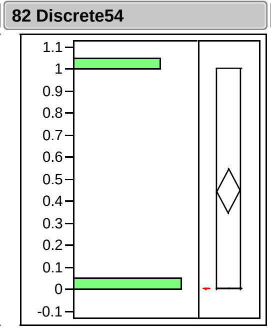
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	0.7500
50.0%	median	0.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.25
Std Dev	0.4351941
Std Err Mean	0.0435194
upper 95% Mean	0.336352
lower 95% Mean	0.163648
N	100

82 Discrete54



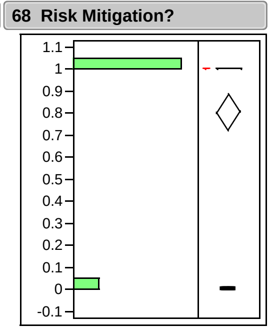
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	0.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.4444444
Std Dev	0.4992206
Std Err Mean	0.0480375
upper 95% Mean	0.5396732
lower 95% Mean	0.3492157
N	108

68 Risk Mitigation?



Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	1.0000
25.0%	quartile	1.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

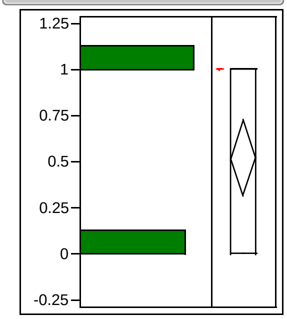
Moments

Mean	0.8039216
Std Dev	0.3989892
Std Err Mean	0.0395058
upper 95% Mean	0.8822905
lower 95% Mean	0.7255527
N	102

Attachment 12

JMP 5.1® Statistical Analysis:
20% Database Variable Distributions (Schedule Cost Category)

82 Discrete54



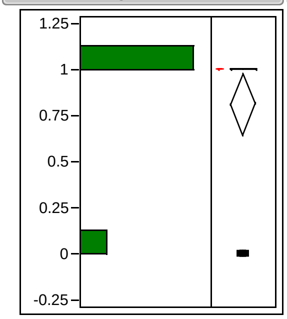
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	1.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.5185185
Std Dev	0.5091751
Std Err Mean	0.0979908
upper 95% Mean	0.7199415
lower 95% Mean	0.3170956
N	27

68 Risk Mitigation?



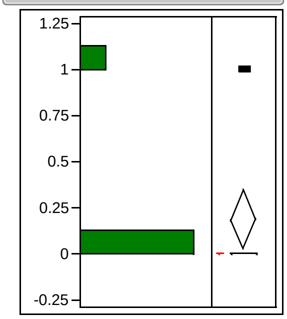
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	1.0000
25.0%	quartile	1.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.8076923
Std Dev	0.4019185
Std Err Mean	0.0788227
upper 95% Mean	0.9700307
lower 95% Mean	0.6453539
N	26

36 MC involvement



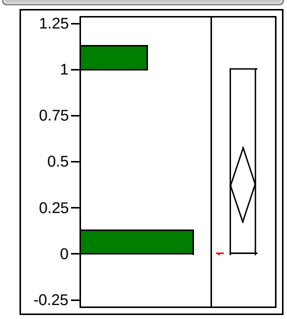
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	0.0000
50.0%	median	0.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.1851852
Std Dev	0.3958474
Std Err Mean	0.0761809
upper 95% Mean	0.3417772
lower 95% Mean	0.0285932
N	27

77 LRIP Planned?



Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	0.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

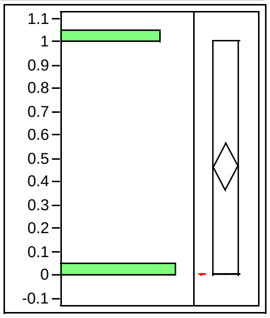
Moments

Mean	0.3703704
Std Dev	0.4921029
Std Err Mean	0.0947052
upper 95% Mean	0.5650398
lower 95% Mean	0.175701
N	27

Attachment 13

JMP 5.1® Statistical Analysis:
80% Database Variable Distributions (Schedule Cost Category)

82 Discrete 54



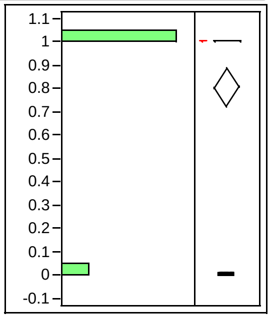
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	0.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.462963
Std Dev	0.500951
Std Err Mean	0.048204
upper 95% Mean	0.5585218
lower 95% Mean	0.3674041
N	108

68 Risk Mitigation?



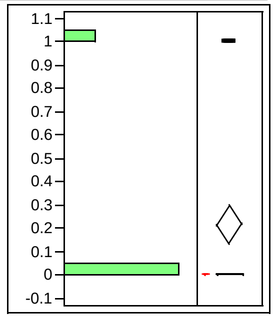
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	1.0000
25.0%	quartile	1.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.8039216
Std Dev	0.3989892
Std Err Mean	0.0395058
upper 95% Mean	0.8822905
lower 95% Mean	0.7255527
N	102

36 MC involvement



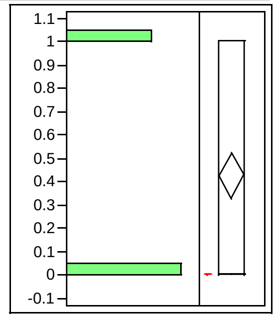
Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	0.0000
50.0%	median	0.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

Mean	0.212963
Std Dev	0.4113103
Std Err Mean	0.0395784
upper 95% Mean	0.2914224
lower 95% Mean	0.1345035
N	108

77 LRIP Planned?



Quantiles

100.0%	maximum	1.0000
99.5%		1.0000
97.5%		1.0000
90.0%		1.0000
75.0%	quartile	1.0000
50.0%	median	0.0000
25.0%	quartile	0.0000
10.0%		0.0000
2.5%		0.0000
0.5%		0.0000
0.0%	minimum	0.0000

Moments

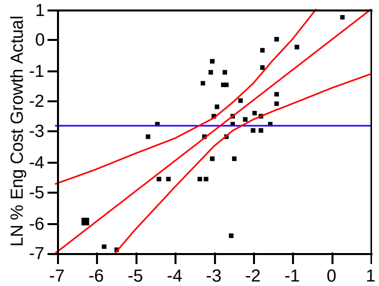
Mean	0.4259259
Std Dev	0.4967879
Std Err Mean	0.0478034
upper 95% Mean	0.5206907
lower 95% Mean	0.3311612
N	108

Attachment 14

JMP 5.1® Statistical Analysis:
Final Full OLS Regression Model (Engineering Cost Category)

Whole Model

Actual by Predicted Plot



LN % Eng Cost Growth Predicted P<.0001
RSq=0.53 RMSE=1.3457

Summary of Fit

RSquare	0.5288
RSquare Adj	0.471684
Root Mean Square Error	1.345684
Mean of Response	-2.77606
Observations (or Sum Wgts)	38

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	67.06337	16.7658	9.2585
Error	33	59.75853	1.8109	Prob > F
C. Total	37	126.82190		<.0001

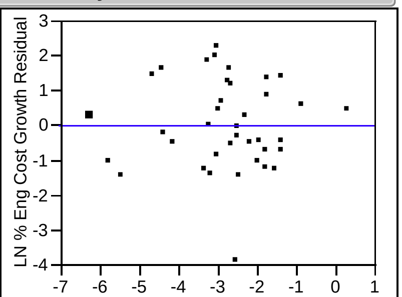
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.305075	1.454603	0.21	0.8352
5 Qty currently estimated for R&D	-0.024198	0.006417	-3.77	0.0006
3 Unit Cost	-0.00262	0.00065	-4.03	0.0003
50 Funding Yrs of Proc Completed	0.1849841	0.050838	3.64	0.0009
57 Maturity of EMD %	-4.439769	1.890935	-2.35	0.0250

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
5 Qty currently estimated for R&D	1	1	25.747116	14.2181	0.0006
3 Unit Cost	1	1	29.439521	16.2572	0.0003
50 Funding Yrs of Proc Completed	1	1	23.976374	13.2403	0.0009
57 Maturity of EMD %	1	1	9.982825	5.5127	0.0250

Residual by Predicted Plot



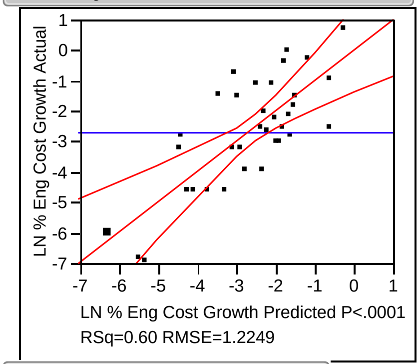
LN % Eng Cost Growth Predicted

Attachment 15

JMP 5.1® Statistical Analysis:
Final Reduced OLS Regression Model (Engineering Cost Category)

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.601516
RSquare Adj	0.548385
Root Mean Square Error	1.224918
Mean of Response	-2.68089
Observations (or Sum Wgts)	35

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	67.94729	16.9868	11.3213
Error	30	45.01274	1.5004	Prob > F
C. Total	34	112.96002		<.0001

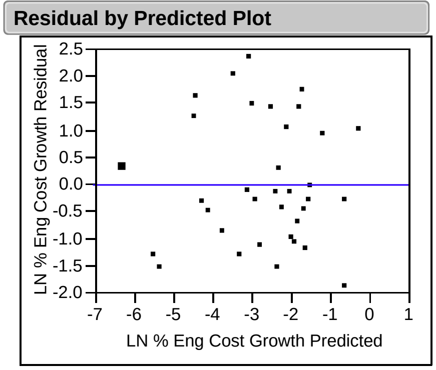
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-2.390668	0.479044	-4.99	<.0001
5 Qty currently estimated for R&D	-0.02506	0.005866	-4.27	0.0002
3 Unit Cost	-0.00295	0.000617	-4.78	<.0001
83 (50/57)	0.1394183	0.039745	3.51	0.0014
76 Program have a MS I?	-1.215384	0.482372	-2.52	0.0173

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
5 Qty currently estimated for R&D	1	1	27.386770	18.2527	0.0002
3 Unit Cost	1	1	34.309851	22.8668	<.0001
83 (50/57)	1	1	18.462740	12.3050	0.0014
76 Program have a MS I?	1	1	9.525277	6.3484	0.0173

Residual by Predicted Plot

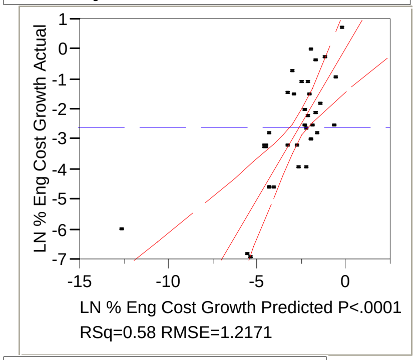


Attachment 16

JMP 5.1® Statistical Analysis:
Final Reduced OLS Regression Model
(Engineering Cost Category, data point 19 excluded)

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.577531
RSquare Adj	0.519259
Root Mean Square Error	1.217059
Mean of Response	-2.58352
Observations (or Sum Wgts)	34

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	58.72204	14.6805	9.9110
Error	29	42.95575	1.4812	Prob > F
C. Total	33	101.67779		<.0001

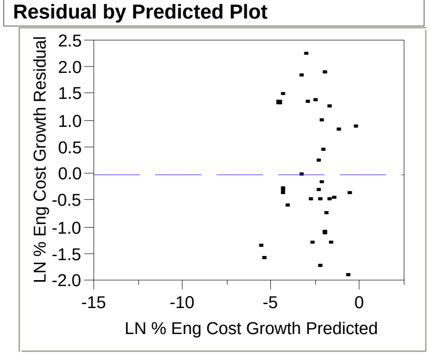
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	-2.29296	0.483138	-4.75	<.0001	.
5 Qty currently estimated for R&D	-0.026373	0.005934	-4.44	0.0001	1.0573576
3 Unit Cost	-0.005825	0.002515	-2.32	0.0278	1.0507865
83 (50/57)	0.137765	0.039515	3.49	0.0016	1.1150278
76 Program have a MS I?	-1.138272	0.483723	-2.35	0.0256	1.1150715

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
5 Qty currently estimated for R&D	1	1	29.261810	19.7550	0.0001
3 Unit Cost	1	1	7.943246	5.3626	0.0278
83 (50/57)	1	1	18.004734	12.1552	0.0016
76 Program have a MS I?	1	1	8.202030	5.5373	0.0256

Residual by Predicted Plot

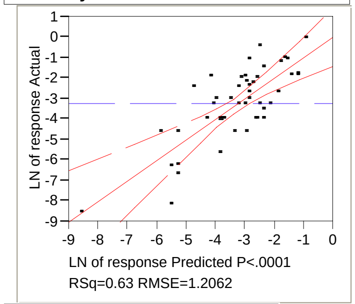


Attachment 17

JMP 5.1® Statistical Analysis:
Final Full OLS Regression Model (Schedule Cost Category)

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.627681
RSquare Adj	0.581141
Root Mean Square Error	1.206221
Mean of Response	-3.2145
Observations (or Sum Wgts)	46

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	98.11548	19.6231	13.4870
Error	40	58.19872	1.4550	Prob > F
C. Total	45	156.31419		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	27	37.166649	1.37654	0.8508
Pure Error	13	21.032068	1.61785	Prob > F
Total Error	40	58.198717		0.6528
				Max RSq
				0.8655

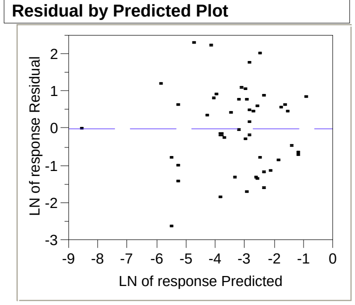
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-9.740498	1.265486	-7.70	<.0001
62 Proc Started based on Funding Yrs?	5.3666135	1.271278	4.22	0.0001
38 Lockheed-Martin	-1.686256	0.468593	-3.60	0.0009
18 Space (RAND)	2.944789	0.981521	3.00	0.0046
50 Funding Yrs of Proc Completed	0.1187131	0.035013	3.39	0.0016
35 N involvement	1.223305	0.382737	3.20	0.0027

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
62 Proc Started based on Funding Yrs?	1	1	25.928248	17.8205	0.0001
38 Lockheed-Martin	1	1	18.841243	12.9496	0.0009
18 Space (RAND)	1	1	13.096719	9.0014	0.0046
50 Funding Yrs of Proc Completed	1	1	16.725871	11.4957	0.0016
35 N involvement	1	1	14.863529	10.2157	0.0027

Residual by Predicted Plot

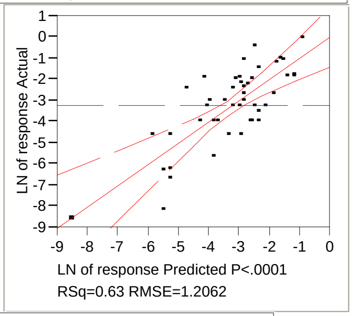


Attachment 18

JMP 5.1® Statistical Analysis:
Final Reduced OLS Regression Model (Schedule Cost Category)

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.627681
RSquare Adj	0.581141
Root Mean Square Error	1.206221
Mean of Response	-3.2145
Observations (or Sum Wgts)	46

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	98.11548	19.6231	13.4870
Error	40	58.19872	1.4550	Prob > F
C. Total	45	156.31419		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	27	37.166649	1.37654	0.8508
Pure Error	13	21.032068	1.61785	Prob > F
Total Error	40	58.198717		0.6528
				Max RSq
				0.8655

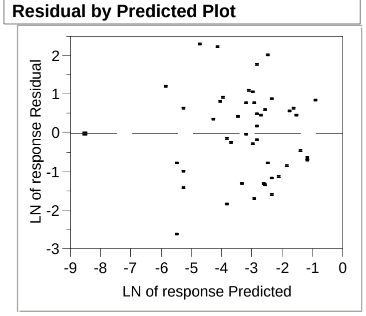
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-9.740498	1.265486	-7.70	<.0001
62 Proc Started based on Funding Yrs?	5.3666135	1.271278	4.22	0.0001
38 Lockheed-Martin	-1.686256	0.468593	-3.60	0.0009
18 Space (RAND)	2.944789	0.981521	3.00	0.0046
50 Funding Yrs of Proc Completed	0.1187131	0.035013	3.39	0.0016
35 N involvement	1.223305	0.382737	3.20	0.0027

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
62 Proc Started based on Funding Yrs?	1	1	25.928248	17.8205	0.0001
38 Lockheed-Martin	1	1	18.841243	12.9496	0.0009
18 Space (RAND)	1	1	13.096719	9.0014	0.0046
50 Funding Yrs of Proc Completed	1	1	16.725871	11.4957	0.0016
35 N involvement	1	1	14.863529	10.2157	0.0027

Residual by Predicted Plot

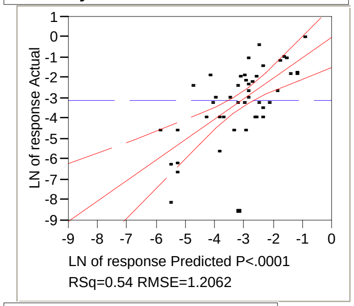


Attachment 19

JMP 5.1® Statistical Analysis:
Final Reduced OLS Regression Model
(Engineering Cost Category, data point 46 excluded)

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.543793
RSquare Adj	0.498172
Root Mean Square Error	1.206221
Mean of Response	-3.09666
Observations (or Sum Wgts)	45

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	69.37201	17.3430	11.9199
Error	40	58.19872	1.4550	Prob > F
C. Total	44	127.57073		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	27	37.166649	1.37654	0.8508
Pure Error	13	21.032068	1.61785	Prob > F
Total Error	40	58.198717		0.6528
				Max RSq
				0.8351

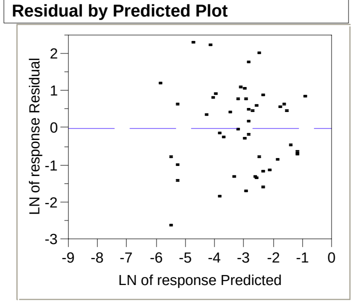
Parameter Estimates

Term		Estimate	Std Error	t Ratio	Prob>|t|
Intercept	Biased	-4.373885	0.432744	-10.11	<.0001
62 Proc Started based on Funding Yrs?	Zeroed	0	0	.	.
38 Lockheed-Martin		-1.686256	0.468593	-3.60	0.0009
18 Space (RAND)		2.944789	0.981521	3.00	0.0046
50 Funding Yrs of Proc Completed		0.1187131	0.035013	3.39	0.0016
35 N involvement		1.223305	0.382737	3.20	0.0027

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F	
62 Proc Started based on Funding Yrs?	1	0	0.000000	.	.	LostDFs
38 Lockheed-Martin	1	1	18.841243	12.9496	0.0009	
18 Space (RAND)	1	1	13.096719	9.0014	0.0046	
50 Funding Yrs of Proc Completed	1	1	16.725871	11.4957	0.0016	
35 N involvement	1	1	14.863529	10.2157	0.0027	

Residual by Predicted Plot



Bibliography

Air Force Materiel Command. *AFMC Financial Management Handbook*. Wright-Patterson AFB OH: HQ AFMC, December 2001.

Bielecki, John V. *Estimating Engineering and Manufacturing Development Cost Risk Using Logistic and Multiple Regression*. MS thesis, AFIT/GAQ/ENC/03-02. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2003 (ADA- 413231).

Calcutt, Harry M. *Cost Growth in DoD Major Programs: A Historical Perspective*. Executive Research Project. The Industrial College of the Armed Forces, National Defense University. Fort McNair Washington, DC 1993.

Cancian, Mark. "Acquisition Reform: It's Not as Easy as it Seems," *Acquisition Review Quarterly*, (Summer 1995).

Department of Defense. *Department of Defense Directive:Tthe Defense Acquisition System*. DoD 5000.1. Washington: GPO, May 2003

Department of Defense. *Department of Defense Instruction Operation of the Defense Acquisition System*. DoD 5000.2. Washington: GPO, May 2003

Department of Defense. "F-22 Fighter Program," *Department of Defense Appropriations Bill*, 2000. http://www.house.gov/appropriations_democrats/view_def_2.htm.

Drezner, J. A., J. M. Jarvaise, R. W. Hess, P. G. Hough, and D. Norton. *An Analysis of Weapon System Cost Growth*. Santa Monica CA: RAND, 1993 (MR-291-AF).

Erwin, Sandra I. "Pentagon's Reform Plans Puzzle Experts." *National Defense Magazine.* November 2002. http://www.nationaldefensemagazine.org/article.cfm?Id=950. 1 Aug 2003.

Grossman, Elaine. "Defense Budgets Include $7.4 Billion Boost For Realistic Costing." Early Bird Article. Inside the Pentagon, February 7, 2002.

JMP® Version 4.0.4, (Academic), CD-ROM. Computer software. SAS Institute Inc., Cary NC, 2001.

JMP® Version 5.0, (Academic), CD-ROM. Computer software. SAS Institute Inc., Cary NC, 2002

McCrillis, John. "Cost Growth of Major Defense Programs." Briefing at the 36th Annual DoD Cost Analysis Symposium. Williamsburg VA. 30 January 2003.

Microsoft® Excel 2000. Version 9.0, IBM, 2.35K, disk. Computer software. Microsoft® Corporation, Redmond WA, 2000.

Moore, Gary W. *Estimating Procurement Cost Growth Using Logistic and Multiple Regression*. MS thesis, AFIT/GAQ/ENC/03-02. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2003 (ADA-413830).

Myers, Dominique. "Acquisition Reform—Inside the Silver Bullet, A Comparative Analysis—JDAM versus F-22," *Acquisition Review Quarterly*, 313-332, (Fall 2002).

Neter, John, Michael H. Kutner, Christopher J. Nachtsheim, and William Wasserman. *Applied Linear Statistical Models*. Boston: McGraw-Hill, 1996.

Office of Management and Budget. *Historical Tables, Budget of the United States Government: Fiscal Year 2004*. Washington: GPO, 2003.

Sipple, Vincent P. *Estimating Engineering Cost Risk Using Logistic and Multiple Regression*. MS thesis, AFIT/GAQ/ENC/02-02. Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2002 (ADA-400576).

Sambur, Marvin Dr., Keys, Ronald E. *Statement Before the Subcommittee on Tactical and Land Forces House Armed Services Committee United States House of Representatives Concerning TACAIR Modernization.* Washington, April 2, 2003.

Suddarth, Steven C. "Solving the Great Air Force Systems Irony," *Aerospace Power Journal*, (Spring 2002).

United States General Accounting Office. *F-22 Aircraft: Issues in Achieving Engineering and Manufacturing Development Goals.* Washington: GPO, March 1999 (GAO/NSIAD-99-55).

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

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
10-03-2004	**Master's Thesis**	Jun 2003 – Mar 2004

4. TITLE AND SUBTITLE	
ESTIMATING COST GROWTH IN ENGINEERING AND SCHEDULE COST CATERGORIES USING A TWO-PRONGED REGRESSION APPROACH	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Chris J. McDaniel, Captain, USAF	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way WPAFB OH 45433-7765	AFIT/GCA/ENC/04-03

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
ASC/FMCE, AFMC Attn: Mr. Michael J. Seibel 1865 4th St Rm 134 DSN: 986-5478 WPAFB OH 45433-7123 e-mail: Michael.Seibel@wpafb.af.mil	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This research confirms the usefulness of the logistic and multiple regression two-step procedure for assessing cost growth in major DoD weapon systems. We compile programmatic data from the Selected Acquisition Reports (SARs) between 1990 and 2002 for programs covering all defense departments. Our analysis concentrates on cost growth in the procurement appropriations of the Engineering and Manufacturing Development phase of acquisitions. We investigate the use of logistic regression in cost growth analysis to predict whether or not cost growth will occur in a program. If applicable, multiple regression is implemented to predict how much cost growth will occur. Our study focuses on the engineering and schedule SAR cost variance categories within the procurement appropriations. We study each of these categories individually for significant cost growth characteristics and develop predictive models for each.

15. SUBJECT TERMS

Logistic Regression, Multiple Regression, Cost Variance, Cost Growth, Selected Acquisition Report, SAR, DoD Cost Growth, Inferential Statistics, Cost Growth in DoD Acquisition Programs, Predicting Cost Growth

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Dr. Edward D. White (ENC)
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	UU	110	(937) 255-3636; e-mail: Edward.White@AFIT.edu