3-2003

Estimating Procurement Cost Growth Using Logistic and Multiple Regression

Gary W. Moore

ESTIMATING PROCUREMENT COST GROWTH USING LOGISTIC AND MULTIPLE REGRESSION

THESIS

Gary W. Moore, Captain, USAF

AFIT/GCA/ENC/03-02

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U. S. Government.

AFIT/GCA/ENC/03-02

ESTIMATING PROCUREMENT COST GROWTH USING LOGISTIC AND MULTIPLE REGRESSION

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 Acquisition Management

Gary W. Moore, BS

Captain, USAF

March 2003

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT/GCA/ENC/03-02

ESTIMATING PROCUREMENT COST GROWTH USING LOGISTIC AND MULTIPLE REGRESSION

Gary W. Moore, BS
Captain, USAF

Approved:

//signed//	
Edward D. White, Major, USAF (Chairman)	date
//signed//	
Michael A. Greiner, Major, USAF (Member)	date
//signed//	
Michael J. Seibel (Member)	date

Acknowledgements

First, I thank my thesis advisor, Maj Edward White, for his support and guidance throughout this research. His leadership was critical to the successful completion of this thesis. Next, I thank my committee members, Mr. Michael Seibel and Maj Michael Greiner. They have been a constant source of support and direction. Further, I would like to thank Capt Vincent Sipple and Capt John Bielecki. Capt Sipple provided the foundation for this research and his knowledge was instrumental throughout. Capt Bielecki provided technical, statistical, and grammatical assistance on weekly basis.

I thank the US Air Force for the opportunities that I have been given. I thank my fellow students for their friendship and support. I thank my wonderful wife whose love, loyalty, and understanding have been instrumental in all that I have accomplished. When times are difficult, she is an endless source of inspiration and strength. I thank God daily for her and for all of the blessings in my life.

Gary W. Moore

Table of Contents

Page

List of Figures

Figure Page

List of Tables

AFIT/GAQ/ENC/03-02

Abstract

Cost Growth in Department of Defense (DoD) major weapon systems has been an on-going problem for more than 30 years. Previous research has demonstrated the use of a two-step logistic and multiple regression methodology to predicting cost growth produces desirable results versus traditional single-step regression. This research effort validates, and further explores the use of a two-step procedure for assessing DoD major weapon system cost growth using historical data.

We compile programmatic data from the Selected Acquisition Reports (SARs) between 1990 and 2001 for programs covering all defense departments. Our analysis concentrates on cost growth in procurement dollar accounts for the Engineering and Manufacturing Development phase of acquisition. We investigate the use of logistic regression in cost growth analysis to predict whether or not procurement cost growth will occur in a program. If applicable, the multiple regression step is implemented to predict how much procurement cost growth will occur. Our study considers all seven SAR categories within the procurement accounts – engineering, schedule, estimating, support, quantity, economic, and other, but we refrain from analyzing these categories individually. Consequently, we focus on the total procurement cost growth incurred from these five categories during the Engineering and Manufacturing Development phase of acquisition.

ESTIMATING PROCUREMENT COST GROWTH USING LOGISTIC AND MULTIPLE REGRESSION

I. Introduction

General Issue

An ongoing problem for over three decades, cost growth in major weapon system acquisitions concerns not only those who work in the acquisition environment, but also the members of Congress and the general public. According to reports by the General Accounting Office, RAND, and the Institute for Defense Analysis, the average cost growth in major DoD acquisition programs ranges anywhere from 20 – 50 % (Calcutt, 1993: i).

Cost growth in major acquisition programs adversely impacts the Defense Department, the defense industry, and the nation. DoD coined the phrase "realistic costing" for the current reform being undertaken in the defense acquisition community. "Under the new costing approach, the Pentagon will adopt program estimates developed by the Cost Accounting Improvement Group (CAIG) in conjunction with a service (Grossman, 2002: 2)." Realistic costing utilizes the CAIG's cost estimating expertise to provide higher quality estimates. DoD's dedication to realistic costing contributed significantly to the cancellation of the Navy Area missile defense program, sending a strong message to the acquisition community. Managers must control their programs, or else. In other words, if managers overrun their budget and breach the Nunn-McCurdy law, their program will be terminated (Grossman, 2002: 2).

For managers to understand and to contain cost growth, they must identify and control the root causes of cost growth. Program managers often resort to a process known as “buffering” in order to increase the accuracy of the baseline estimate and to limit the programs likelihood of incurring cost overruns; this process necessitates that the manager accurately identify risks related to potential cost growth in program estimates and assign appropriate dollar values to these risks. While ultimately responsible for their programs, managers require support from the cost estimating community, the contracting office, and the defense contractor. Specifically, management relies on the cost estimating community to assign accurate dollar values to specific risk factors and include these dollar amounts into the cost estimate.

Specific Issue

Cost estimators utilize a vast assortment of methods to determine and assign dollar amounts to specific risk factors. Oftentimes, cost estimators rely on subjective means, such as expert opinions, for making these dollar assignments. When available, estimators normally opt for more objective methods, such as gathering historical data and comparing analogous systems. Analysts prefer historical data when available, because in the past, it has provided more accurate estimates and it requires the analyst to understand the relationships between program attributes and observed cost increases. When possible, the analyst should group historical cost growth data into different categories and then analyze these categories to determine if different types of cost growth have different and distinct predictors. Statistical regression techniques prove useful for determining such relationships. Thus, this research utilizes statistical regression to find predictors of cost growth (Sipple, 2002: 2).

Scope and Limitations of the Study

A key aspect of any discussion of DoD cost growth is the Selected Acquisition Report (SAR). Since 1969, Congress has required DoD to submit SARs on its major acquisition programs. SARs contain information necessary to identify the three cost estimates, planning, development, and current, which are useful in analyzing program cost growth (Calcutt, 1993: 3). The Planning Estimate is developed during the Concept Exploration and Definition phase of the acquisition cycle. The Development Estimate is established at Milestone II, which is the beginning of Manufacturing Development. The Current Estimate is the most up-to-date estimate as to what the program will cost at completion.

When determining cost growth, the Government Accounting Office compares the Planning Estimate to the Current Estimate, while the Institute for Defense Analysis and RAND compare the Development Estimate to the Current Estimate. These different interpretations of cost growth are a matter of the investigating organization's purpose. The Government Accounting Office is interested in providing Congress a top-level review of DoD's ability to plan and manage acquisition programs. RAND and the Institute for Defense Analysis are concerned with understanding the factors that cause cost growth and developing a formula to account for these factors (Calcutt, 1993: 7). Since our research focuses on the factors that cause cost growth, we define cost variance as the difference between the Current Estimate to the Development Estimate and cost growth as positive cost variance.

The SARs separate program cost variance into seven categories: Economic, Quantity, Estimating, Engineering, Schedule, Support, and Other (Calcutt, 1993: 4). The division of cost growth into separate components enables us to perform standardized comparisons of variances across acquisition programs. Due to the level of detail available in the SARs and the ease of accessibility, we utilize SAR data in our search for predictors of cost growth. Previous researchers constructed a database from the SARs that is both accurate and relevant to our research. We employ the aforementioned database in our research efforts. Additionally, we update the research database to include only the most recent SAR for each program.

In this study, we scope our research to include only programs that use the Development Estimate as the baseline estimate and by considering only the most recent SAR available for each program. We measure cost growth as a percentage increase in cost between the Current Estimate and the Development Estimate as recorded in the SAR. Further, we limit our study to procurement cost growth that occurs in the EMD phase of the acquisition life cycle. Additionally, complications arise from utilizing the SAR database that further limits our research (e.g. security classification, etc.). We discuss many of these limitations in depth in Chapter III. Lastly, baseline estimates oftentimes include unknown budgeted amounts for risks; these "buffers" further limit the interpretation of the results of this research.

In our study, we use historical data to help us identify candidate predictor variables for cost growth; this inferential study differs from a majority of the past DoD studies on cost growth. Most of the previous research on cost growth within DoD utilized descriptive methods. In part, DoD preferred descriptive studies because of its focus on macro-level cost growth. We find only a few historical studies that apply multiple regression and even fewer studies consider using logistic regression techniques. More specifically, our study has only one predecessor to date; Sipple (2002) provides the framework and methodology for this research effort. Our study builds on Sipple's research and mirrors his research in nearly every aspect with the only exceptions being the area of application and the depth of analysis.

Research Objectives

This study has three main objectives. First, we utilize logistic regression to determine if certain program characteristics predict whether a program experiences procurement cost growth during the EMD phase of the acquisition cycle. "Logistic regression differs from multiple regression in that it predicts a binary response. In our case, the binary response is: Does a program experience cost growth, Yes or No (Sipple, 2002: 5)?" Once we establish that a program will experience procurement cost growth, we then use multiple regression to determine if certain program characteristics predict the amount of procurement cost growth in the EMD phase of development. Finally, we seek to discover the nature of these predictive relationships. We then develop cost-estimating relationships for predicting whether a program will have procurement cost growth and for predicting the amount of cost growth the program will incur during the EMD phase of development. For predicting the amount of procurement cost growth, we develop cost-estimating relationships that return point estimates and that provide the estimator with an upper bound based on a specified level of confidence. We discuss confidence bounds further in Chapter IV.

Synopsis of Research

This study attempts to leverage off past cost growth research to create more accurate models for the financial management community, so they may better estimate risk in dollar terms according to program characteristics. To develop these models we perform logistic and multiple regression on data from the SAR database. This study concentrates on procurement cost growth during the EMD phase of the acquisition cycle. We limit our analysis to include only programs recorded in the SAR database over the last decade. Further, for purposes of this research, we measure cost growth as a percentage increase in cost between the Current Estimate and the Development Estimate. Finally, while program managers must choose an avenue for handling the issues associated with cost growth, this research attempts to reduce measured cost growth by helping cost estimators predict cost growth earlier in the program and with a greater level of accuracy.

II. Literature Review

Chapter Overview

This chapter provides the background information necessary to facilitate the research carried forth in this thesis. We first describe the overall cost-estimating environment and follow with details of historical studies pertaining to cost growth. Lastly, we confine the bulk of our literature review to focus on the Sipple Study (2002). Specifically, we concentrate on this study because it establishes a database, predictor variables, and a methodology from which we base our research efforts.

Cost Estimating

The DoD cost estimating community considers cost risk as the "funds set aside to cover projected cost growth," and it defines cost growth as the "increase in the cost of a system from inception to completion." Thus, cost risk represents the predicted dollar amounts associated with a program and cost growth represents the actual incurred dollar amounts of a program (Coleman, 2000:3). The research efforts carried forward in this thesis serve to minimize the effects of cost growth by providing the estimator an invaluable tool for assessing cost risk.

The *AFMC Financial Management Handbook* gives the Air Force perspective on risk analysis:

> Cost estimating deals with uncertainty. The analyst attempts to describe in the best terms possible the probability distribution of a cost event in the future. One value for the cost estimate is the result of one prediction of that future event. Risk Analysis is a careful consideration of the areas of

> uncertainty associated with future events. The preferred method is to identify the risk in the program and then quantify it into dollars. *(AFMC Financial Management Handbook*, 2001:11-12)

Thus, the cost analyst estimates risk in terms of dollars and establishes a probability distribution to express the range of possible outcomes and their probabilities. The handbook distinguishes program risk as "the uncertainties and consequences of future events that may affect a program" (*AFMC Financial Management Handbook*, 2001:11-12). The services use logical methods to assess risk in different areas of a program and then quantify that risk within their estimates. The estimator quantifies the risk in these areas, but ultimately, the program manager decides which dollar amounts to incorporate in the final estimate (Sipple, 2002: 13).

The military cost estimator employs a multitude of different risk assessment techniques when performing an estimate. The method chosen depends on the type of risk estimated, the level of detail needed in the estimate, the level of accuracy required in the estimate, the timeframe within which the estimator has to complete the estimate, the skill of the estimator, the data and tools available to the estimator, and any office policies directing estimating practices (Sipple, 2002: 14).

Figure 1 shows the risk assessment techniques recognized by the Ballistic Missile Defense Organization (BMDO) cost-estimating community. In addition, this chart shows the relationship between the level of accuracy required in the estimate, the time required to complete the estimate, the difficulty associated with performing the estimate, and the type of estimate. The "degree of precision" needed in an estimate drives the type of estimate used: as the degree of precision needed increases, the estimate techniques used become more detailed and difficult (Coleman, 2000:4). It should be noted, that these techniques are widely used throughout the Defense acquisition community.

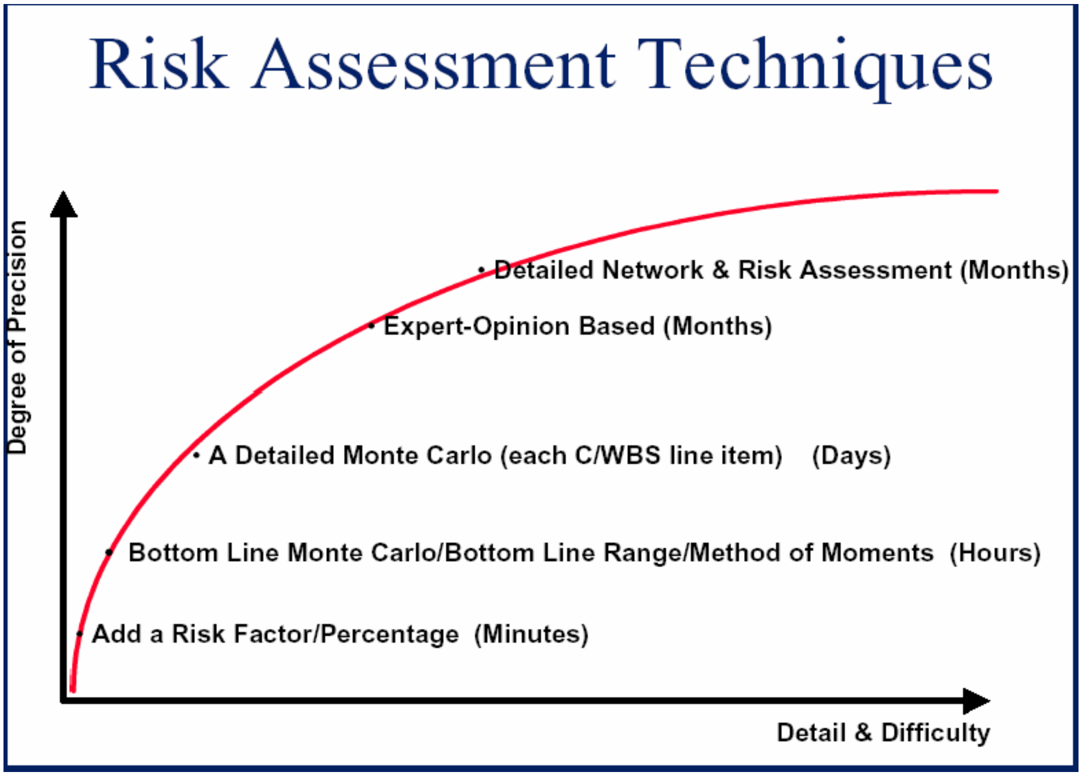


Figure 1. Risk Assessment Techniques (Coleman, 2000:4-9)

Starting with the most accurate, but also the most difficult and time-intensive method, the Detailed Network & Risk Assessment technique requires an extraordinary amount of effort to complete due to the meticulous detail required in its schedule and task breakout. This method assigns probability distributions to the schedule item durations to create a stochastic model from which schedule slip can be estimated. The estimator analyzes this information using Monte Carlo Simulation to estimate the cost. The drawback of using such a technique is that it cannot be simplified or condensed and still be accurate (Coleman, 2000: 9).

The technique representing the next lower level of precision is the Expert-Opinion-Based technique. This method surveys technical experts to determine the probable distributions of Work Breakdown Structure (WBS) item costs and then incorporates the use of Monte Carlo Simulation to provide a range of possible costs. Ultimately, the accuracy of this method depends on the abilities of the technical experts to evaluate these costs in light of their past experiences. Invariably, the problem with this technique is that technical experts are not always cost experts and may not have a real sense of how much things cost, or how much costs can grow (Coleman, 2000: 12).

Monte Carlo Simulation, although less precise than the previous two methods, is one of the most commonly used techniques for estimating uncertainty. This method employs running a simulation for "each C/WBS line item," where C/WBS is the Cost or Work Breakdown Structure of the program. Although the two previous methods incorporate Monte Carlo Simulation in their assessment, this method differs because it develops probability distributions of cost outcomes based on historical databases instead of lengthy surveys or program evaluation and review technique analyses. The weakness of this technique lies in the accuracy, applicability, and currency of the data compiled in the database (Coleman, 2000: 17). Despite this weakness, Monte Carlo Simulation arguably provides the most "bang for the buck," since it produces a reasonable amount of accuracy for the time that the analyst puts into it (Coleman, 2000: 4).

The Bottom Line Monte Carlo, Bottom Line Range, and Method of Moments techniques represent the next lower level of precision and detail. These methods might use analogous system methodology or a limited database to construct an estimate, or they might rely on expert opinion to develop an estimate. Oftentimes, these methods utilize Monte Carlo Simulation techniques, but they focus on higher levels of the WBS (Sipple, 2002: 18).

The least detailed and least accurate of the risk assessment methods is "Add a Risk Factor/Percentage." This technique relies solely on technical expert judgment to assign a high-level, subjective risk factor for the estimate (Coleman, 2000: 4).

Past Research in Cost Growth

Before analyzing data, we consider logical relationships in the acquisition environment that might explain cost growth. Previous research facilitates our search for possible predictor variables and ultimately a formula to accurately forecast cost growth. In this section, we list various historical studies that address cost growth. Two important factors should be noted when reviewing the historical research. First, a majority of the cost growth research does not break cost growth down into its components. Second, the bulk of these studies do not partition cost growth into separate phases of development. Therefore, one cannot directly tie the results from any of the studies considered directly to the nature of procurement cost growth during EMD.

Sipple (2002) conducts an exhaustive review of all cost growth studies performed during the past ten years. From this review, Sipple gains valuable insight into the root causes of cost growth. Additionally, he finds extensive amounts of research devoted to establishing predictive relationships and determining predictor variables. Our search for new or additional relevant cost growth research produces no results. So for purposes of this research, we utilize the findings from Sipple's literature review as a foundation. For a comprehensive review of the cost growth studies listed below in Table 1 refer to Sipple (2002).

Table 1. Historical Cost Growth Studies (Sipple, 2002)

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

Sipple collects data from the SAR database on 115 major acquisition programs, spanning the years from 1990 to 2000. Sipple then assembles an extensive database with over 70 possible predictor variables. From this database, he constructs a plethora of regression models aimed at predicting EMD cost growth directly related to engineering changes. Sipple finds that a two model system utilizing logistic and multiple regression techniques most accurately represents the projected cost growth without violating the underlying regression assumptions (Sipple, 2002: 125).

Sipple develops the following models: Model A utilizes logistic regression techniques to predict whether a program will have cost growth or not. Model B employs multiple regression techniques to forecast the amount of cost growth a program will have once model A deems that the program will incur cost growth. Sipple conducts validation testing on model A and determines that the model predicts accurately 69% of the time. Sipple performs further validation testing on model B and determines that at an 80% confidence level, this model accurately projects the amount of cost growth 69% of the time. Model A utilizes seven variables from the list of plausible predictor variables, while model B only incorporates three of the 78 possible predictor variables. Table 2 and Table 3 show the predictor variables and their associated *p*-values for models A and B, respectively (Sipple, 2002:122).

Table 2. Predictors Variables for Model A

Predictor	*P*-value
RAND Modification	0.0037
Actual Length of EMD (MSIII-MSII in mos)	0.0029
Length of R&D in Funding Yrs	0.0020
MSIII-based Maturity of EMD %	0.0148
Length of Prod in Funding Yrs	0.0012
Actual Length of EMD (using IOC-MSII in mos)	0.0154
Land Vehicle	0.0132

Table 3. Predictor Variables for Model B

Predictor	*P*-value
Maturity from MSII (in mos)	0.0069
No Maj Def KTR	0.0024
PAUC	0.0410

Sipple's study not only develops the list of plausible predictor variables for this research, but also establishes the use of both logistic and multiple regression techniques for determining cost growth. Additionally, Sipple specifically concentrates on cost growth during EMD instead of focusing on overall program cost growth. While Sipple does not consider procurement cost growth in the EMD phase, he does breakdown the previous barrier of only considering macro-level cost growth. Sipple's research provides the predictor variables, the methodology, and the framework necessary to pursue this study.

Purpose of Review

In this chapter, we reference many historical studies that investigate a multitude of different databases using a variety of statistical techniques in the quest to explain cost growth in DoD acquisitions. The Sipple study (2002) establishes a general list of possible predictor variables that are ascertainable within the SAR database. We provide a complete list of these predictor variables in Chapter III. Additionally, we implement the research database and methodology founded by Sipple (2002) to provide the framework for our research efforts. None of the aforementioned studies deals directly with procurement cost growth in the EMD phase, but the results from these studies provide the insight necessary to successfully find predictors of procurement cost growth in the current study.

III. Methodology

Chapter Overview

This chapter outlines the statistical procedures carried forth in this research. We first assess our use of the SAR and explain the process of data collection and database construction. Secondly, we discuss Sipple (2002) in depth since it serves as the cornerstone for our research efforts. The purpose of this discussion is two fold. To begin with, it provides insight into possible predictive relationships for determining cost growth. Additionally, this literature serves as a foundation of knowledge from which we further analyze the results of this research. Lastly, we describe the exploratory data analysis and regression techniques that we use.

Data Assessment

We use the Selected Acquisition Report (SAR) database as the sole source for cost variances and other information included in this analysis. The SAR provides cost variance data in both base year and then year dollars. We do not include estimated inflationary effects in our analysis, therefore, we use cost variances reported in base year 2001 dollars for analysis. Furthermore, this format facilitates conversion of the various base years of individual estimates into a single constant year, making comparison across programs more feasible. Lastly, the SAR records cost variances in seven different categories:

- Economic: changes in price levels due to the state of the national economy

- Quantity: changes in the number of units procured
- Estimating: changes due to refinement of estimates
- Engineering: changes due to physical alteration
- Schedule: changes due to program slip/acceleration
- Support: changes associated with support equipment
- Other: changes due to unforeseen events (Drezner, 1993:7)

Sipple's study in 2002 analyzes cost variance during the EMD phase of development due specifically to engineering changes. This thesis focuses on procurement cost variance during the EMD phase, but does not specifically target any one category. We do not target a specific category because the cost estimator is only concerned with total cost growth. Further, we do not target total cost growth in its aggregate form as past research has shown it to be unpredictable at that level. Therefore, we focus our efforts on the predicting cost growth at the next logical level. This entails choosing a type of funding (i.e. procurement) and limiting the study to one phase within the acquisition life cycle (i.e. EMD).

The SAR database contains historical, schedule, cost, budget, and performance information for major acquisition programs from all military services. The SAR database contains files on only ACAT IC and D programs (Knoche, 2001:1). Therefore, the programs listed in the SAR consistently represent programs with high-level government interest. For security reasons, we do not use any information from the SAR that has a security classification in the compilation of our database.   Thus, the database we construct for this research represents an assemblage of the programmatic details of some of the most important DoD programs, but this database is not all-inclusive.

Previous studies establish the use of SAR data in cost-growth research. In the early 1990's, the researchers at RAND modify and compile selected information from individual SARs in spreadsheet format. Unfortunately, the RAND database does not segregate cost growth into the seven SAR categories. Furthermore, the most recent entries in the RAND database date back to the early 1990s. These shortcomings limit the use of the RAND database in our research efforts.

In 2002, Sipple researches the SAR and compiles a modified database. Similar to the RAND study, Sipple constructs a database containing selected information from individual SARs. Contrary to the RAND study, Sipple breaks cost growth into the seven categories listed above. In addition, Sipple's database contains SAR data entries as recent as the year 2000 (Sipple, 2002: 49). Therefore, Sipple's database serves as the foundation for our data collection efforts. Additionally, our effort updates this database to include the year 2001.

The SAR Database as a Research Tool

According to Calcutt, a key aspect of any discussion of cost growth is the SAR (Calcutt, 1993: 3). Calcutt notes that the SAR data is imperfect due to several factors, but the SAR is the most convenient source of data for studying cost growth. In a 2002 study, Sipple conducts an extensive search of previous SAR related cost-growth research. Sipple's review further confirms that there are limitations associated with using the SAR to study cost growth. Additionally, he finds that the SAR is the logical choice from which to calculate cost-growth research, regardless of its shortcomings. Sipple provides a complete analysis of the SAR limitations, but the following list contains only those problems most prevalent (Sipple, 2002: 49-56):

- 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 (Hough, 1992:v)

Data Collection

The SAR database contains an overwhelming amount of information that proves useful for the research conducted herein. The SAR covers a broad spectrum of programs and contains reports from all of the services. Thus the SAR contains thousands of individual reports and each report contains a plethora of information on that particular program. Data collection involves "scrubbing" the database to determine which data is most pertinent to our research efforts. Additionally, the SAR database is lacking information for some programs and completely excludes other programs (when the entire file is classified). This research does not further restrict the information provided in the SAR; the data collection effort only excludes data that has a security classification.

In Sipple's research efforts, he constructs a database that contains SAR data from 1990 through the summer of 2000. To ensure validity and reliability, this data analysis requires the most current information to capture recent trends. Therefore, we begin our data collection with the most recent SARs available. Specifically, the latest SARs at our disposal are from December 2001. Thus, our data collection efforts begin with those SARs and work backwards through the summer of 2000. These reports are then incorporated into the previous database constructed by Sipple.

Consequently, the database now spans from 1990 through 2001. We do not complete data collection by merely selecting which SAR reports to incorporate into our analysis. Sipple explains that once we select files for collection, we must determine what information within the files will prove useful for predicting procurement cost growth. Furthermore, we must determine what form of the data will prove most useful in this analysis. Additionally, for programs that have more than one SAR available we utilize only the latest SAR; this ensures that we have independence of data points (Sipple, 2002: 58).

Constructing a Foundation

A 2002 SAR-based cost-growth study by Vincent Sipple serves as the cornerstone for the literature basis. Sipple conducts both logistic and multiple regression analyses on RDT&E cost growth due to engineering changes. From this study, we form general impressions about cost growth as it relates to different programmatic characteristics. We go on to investigate several previous studies pertaining to cost growth and risk analysis, but we fail to find a study that shares the exact focus as our study – procurement cost growth during the EMD phase. Much of this research mirrors the efforts carried forth in Sipple's study, but our study differs from Sipple's in one very important way: Sipple's study focuses on the ability of candidate predictors to predict one of the SAR categories of cost growth, specifically, "Engineering changes." Our study analyzes the ability of predictor variables to predict total procurement cost growth in the EMD phase. Even with these differences, the literature review still provides useful insight toward our purpose. Like Sipple, we limit ourselves to predictors that we find within the SAR data, so some of the clues established in previous research will not be explored further in this study and remain fertile ground left for future researchers to explore (Sipple, 2002: 45).

Search for Predictors of Cost Growth

From Sipple's research, we identify possible cost-growth predictor variables for the research efforts carried forth in this thesis. We expand our search to contain not only known or logical predictors of cost growth, but also to include any variable that we suspect has a possible predictive relationship. We then narrow our search to focus only on variables that the cost analyst either knows or is able to estimate at the time the program office accomplishes the Development Estimate. If the cost analyst has no idea of the value of a predictor variable at the time he produces the estimate, then the variable is insignificant regardless of how accurately it predicts cost growth, because the analyst is unable to use it to produce a cost estimate of the response variable. Thus, the model that we produce does not include such esoteric variables (Sipple, 2002: 47).

Finally, we must ensure that that cost analyst understands the relationship between the predictor variables and the response variables in any models we discover. "If the estimator does not understand the variables, two problems may arise. First, the estimator might lack faith in the model, causing him to discredit its results. Second, even if the estimator supports the model, he will not have the ability to support it in the event it falls under management scrutiny (Sipple, 2002: 48)." Thus, we determine that the predictors we employ in our model must satisfy two conditions. First, although these variables do not have to demonstrate a direct causal relationship with the response variable, they must have some logical tie to the response variable that the estimator can easily understand. Second, any predictor variables we unearth must be available to the estimator at the time of the Development Estimate (Sipple, 2002: 48).

Exploratory Data Analysis

For the results of our research to be valid, we must ensure that the techniques we utilize are employed correctly. A basic assumption of multiple regression requires that the response variable be from a continuous distribution. A review of Sipple's study, indicates that *Engineering* cost growth during EMD is from a mixed distribution. About half of the distribution is continuous, while the other half is massed around zero. In addition to the mixed distribution, Sipple finds that a few of the programs have negative cost variance. When we perform a preliminary analysis of our data, we find that the distribution for procurement cost growth during EMD exhibits identical characteristics to those found by Sipple. Therefore, we duplicate the procedures Sipple established in his research, making slight modifications where necessary. An overview of these procedures follows:

We first split the data into discrete and continuous distributions. We then utilize logistic regression to analyze the discrete distribution and multiple regression to analyze the continuous distribution. Thus, we develop two models: a logistic regression model that analyzes the full data set to predict whether or not a program will have procurement cost growth, and a multiple regression model that analyzes only programs containing procurement cost growth to predict the amount of cost growth we expect. For the logistic regression portion of our analysis, we convert all negative cost growth to zero cost growth. Furthermore, to ensure that we construct a robust model, we set approximately 20 percent of our data aside for validation before we begin any regression analysis. We use the "column shuffle" command in JMP® 4 (SAS Institute, 2001) to randomly select which data we set aside. Finally, before performing regression, we must also choose the response and candidate predictor variables (Sipple, 2002: 59).

Response Variables

As mentioned in Chapter I, this research seeks to find predictors of procurement cost growth in the EMD phase of development. We concern ourselves with two different response variables, one that indicates if procurement cost growth will occur and another that expresses the degree to which procurement cost growth occurs. The first of the two, we express as a binary variable where the value '1' means that we estimate a program will have cost growth in procurement dollars, while the value '0' means that it will not. We call this variable *Procurement Cost Growth.*

In order to construct the most useful model possible, we decide that the second response variable should be the percentage of procurement cost growth. This format applies equally well to programs with both large and small acquisition costs, whereas the dollar amount format requires us to consider the effects of program size on the model for the results to intuitively make sense. For example, a model with length of procurement and maturity from milestone III decision might produce a predicted cost growth of 20 million dollars for both a 20 million dollar program and a 2 billion dollar program. Although, these results may prove statistically valid, they would be difficult for decision maker to put into context (Sipple, 2002: 60). Thus, we strive to find a model to predict percent change in procurement cost and therefore, we use the Development Estimate as the denominator of the percentage. We call this second response variable *Procurement Cost Growth %.*

Predictor Variables

A plethora of possible predictor variables exist within the SAR data. We wish to create a tool that enables cost estimators to develop more accurate estimates, so the inputs (predictor variables) for such a tool must be obtainable at the time of the estimate. However, we do not exclude variables from our analysis solely based on this availability criterion. Instead, we evaluate those variables to determine if predictive capabilities exist with the hope of finding some correlated variable that is available to the estimator (Sipple, 2002: 61).

The predictor variables we attain from the SAR fall into five categories: program size, physical type of program, management characteristics, schedule characteristics, and other characteristics. We create two subcategories within each main category and classify the variables as either "binary" or "continuous" variables. We provide a list of the predictor variables below; this list is sorted by category and then by subcategory. Sipple provides a brief description of the subcategories that includes explanation of ambiguous elements where necessary (Sipple, 2002: 61):

Program Size Variables

- Total Cost CY $M 2002 (Continuous)
- Total Quantity (Continuous)
- Prog Acq Unit Cost (Continuous)
- Qty during PE (Continuous)
- Qty planned for R&D$ (Continuous)

Physical Type of Program

- Domain of Operation Variables (Binary)
    - Air, Land, Space, & Sea
- Function Variables (Binary)
    - Electronic, Helo, Missile, Aircraft, Munition, Land Vehicle, Ship, & Other

Management Characteristics

- Military Service Management (Binary)
    - Services (Svs) >1, Svs >2, Svs>3, Service = Navy Only, Service = Joint, Service = AF Only, Lead Svc = Army, Lead Svc = Navy, Lead Svc = DoD, Lead Svc = AF, AF Involvement, N Involvement, MC Involvement, & AR Involvement
- Contractor Characteristics (Binary)
    - Lockheed-Martin, Northrup Grumman, Boeing, Raytheon, Litton, General Dynamics, No Major Defense KTR, More than 1 Major Defense KTR, & Fixed-Price EMD Contract

Schedule Characteristics

- RDT&E and Procurement Maturity Measures (Continuous)
    - Maturity (Funding Yrs complete), Funding YR Total Program Length, Funding Yrs of R&D Completed, Funding Yrs of Prod Completed, Length of Prod in Funding Yrs, Length of R&D in Funding Yrs, R&D Funding Yr Maturity %, Proc Funding Yr Maturity %, & Total Funding Yr Maturity %
- EMD Maturity Measures (Continuous)
    - Maturity (Funding Yrs complete), Funding YR Total Program Length, Funding Yrs of R&D Completed, Funding Yrs of Prod

Completed, Length of Prod in Funding Yrs, Length of R&D in Funding Yrs, R&D Funding Yr Maturity %, Proc Funding Yr Maturity %, & Total Funding Yr Maturity, Maturity from MS II in mos, Actual Length of EMD, MS II-based Maturity of EMD%, IOC-based Maturity of EMD%, & FUE-based Maturity of EMD%

- Concurrency Indicators (Binary & Continuous)
  - MS III Complete, Proc Started based on Funding Yrs, & Proc Funding before MSIII (Binary)
  - Concurrency Measure Interval & Concurrency Measure % (Continuous)

Other Characteristics

- # Product Variants in this SAR (Continuous)
- Security Classification (Binary) – Class S, Class C, Class U, & Class at Least S
- Risk Mitigation (Binary)
- Versions Previous to SAR (Binary)
- Modification (Binary)
- Prototype (Binary)
- Dem/Val Prototype (Binary)
- EMD Prototype (Binary)
- Did it have a PE (Binary)
- Significant pre-EMD activity immediately prior to current version (Binary)
- Did it have a MS I (Binary)
- Terminated (Binary)

The multitude of defense contractors available presents a significant quandary to our use of defense contractor as a predictor variable, because it decreases the likelihood of repeat contractors on different programs. The small number of repeat contractors makes it difficult to obtain statistically relevant results. Fortunately, in the 1990s, the defense industry was marked with an intense movement towards contractor consolidation. Sipple notes these consolidations and provides a complete list of the contractor consolidations that occur from 1993-2000. We defer to Sipple's consolidation of contractors as it provides us sufficient data points for most of the categories to achieve useable results from the regressions (Sipple, 2002: 67). Table 4 provides a complete list of the new contractor variables as determined from Sipple's study.

Table 4. Contractor Variables

New Consolidated Contractor Variables
Boeing
General Dynamics
Litton
Lockheed-Martin
Northrop Grumman
Raytheon
No Major Defense Contractor
More than 1 Major Defense Contractor

The EMD maturity variables present additional quandaries. We address issues of ambiguity and scarcity within the underlying schedule parameters. Sipple determines that "MS II and MS III dates often have different versions of the same schedule item, making unclear which date to use for computation. In order to capture the entire EMD effort, we use the earliest MS II date and the latest MS III date available for our maturity calculations (Sipple, 2002: 66)." Like Sipple, we determine that when EMD maturity variables use Initial Operational Capability (IOC) or First Unit Equipped (FUE) computations are incorporated into our model, we face a scarcity of data points, thus limiting the potential use of these as predictors in our regression models.

Logistic Regression

Logistic regression is normally used to analyze possible predictive relationships when the response is either nominal or ordinal. Logistic regression mainly predicts binary outcomes, usually coded '0' and '1' (Neter, 1996:567). We utilize logistic regression to develop a model that predicts whether a program will have procurement cost growth or not. Therefore, in our database, we code each program that incurs cost growth with a '1' and each program that has either no cost or negative cost growth with a '0.' Since an estimator would not assess negative cost growth in an estimate, we do not consider negative cost growth in our model. Our cost-growth distribution contains only 0's and 1's, thus we characterize *whether or not a program has procurement cost growth* as a Bernoulli random variable with probability *p* of success (success=1) (Neter, 1996:568).

Sipple established the following guidelines for utilizing logistic regression in cost-growth analysis:

> We use JMP® 4 (SAS Institute, 2001) software to accomplish the logistic regression in order to help us identify the best model for estimating whether or not a program will have cost growth. JMP® uses maximum likelihood to estimate the coefficients of our model. Because JMP® has no automatic method, such as stepwise, for logistic regression, we manually compute thousands of individual regressions, recording our results on spreadsheets. We start with one-predictor models of all possible variables. Then we regress using all combinations of two-predictor models and record the results. We continue this process, eventually whittling down the best combinations for use at the next level in order to cut down on the amount of regressions necessary. We stop when we reach a model for which the gain of adding another variable does not warrant the additional complexity of the model that another variable adds. We intend to find several candidate models for each number of predictors and then narrow down to the best one for each number of predictors and validate the model using about 20 percent of the data that we set aside for validation (Sipple 2001: 70).

Multiple Regression

We utilize multiple regression to discover prediction models for the percent of procurement cost growth based on more than one predictor variable. Our efforts during multiple regression focus not only on individual variables, but also include logical interactions between variables that may enhance their predictive relationships. We present the following simplistic fictional scenario (Figure 2). If the interactions are not considered, then the center line shows the amount of cost growth (30%) associated with aircraft type across all services considered. When we consider interactions, we find that the cost growth varies depending on both the aircraft type and the lead service involved (i.e. 40 percent for Air Force helicopters).

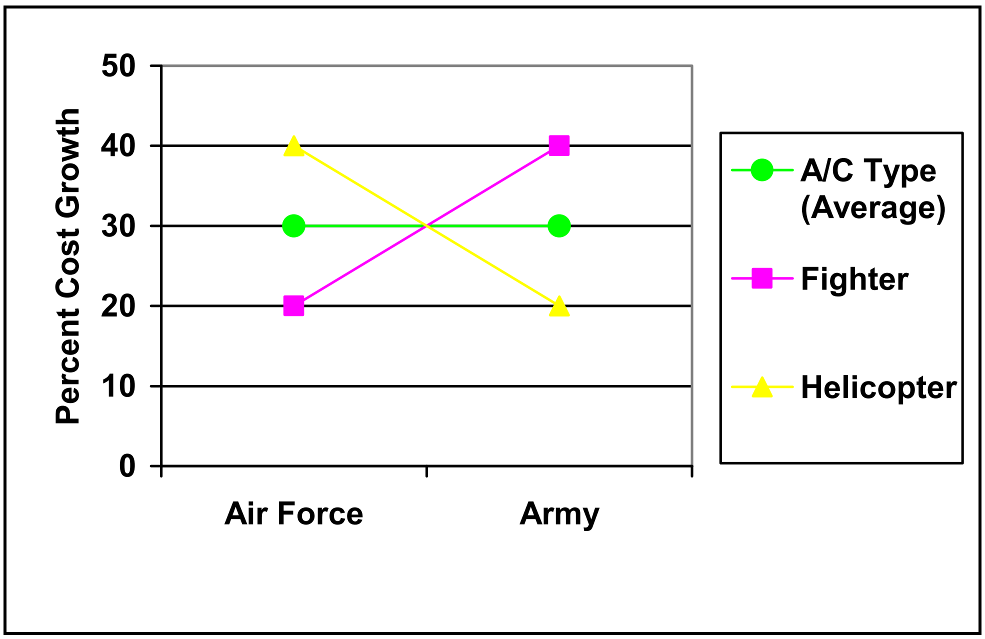


Figure 2. Explanation of Interacting terms

As with logistic regression, we utilize the following multiple regression guidelines established by Sipple in his research efforts:

> We use JMP® for the multiple regression analysis. We use the stepwise method to identify those predictor variables that have a statistically significant impact on the ability of the model to predict our response variable, *Engineering %*.

> From our stepwise analysis, we build models using the standard least squares method, whereby JMP® estimates the form of the functional relationship between the predictors and the response variable that minimize the sum of squared deviations from the predicted values at each level of the predictors (Neter, 1996).
>
> Because of the large amount of candidate predictor variables, we exceed JMP®'s stepwise calculation abilities when we include all of our variables in a single run. In addition, we seek models with varying numbers of predictors. Thus, we must repeat the stepwise and standard least squares several times in order to achieve the desired results. As with logistic regression, we discover several candidate models for each number of predictors. Then we narrow our results to the best model for each number of predictors. We continue adding variables to the model until the number of variables equals about one tenth of the number of data points used in the model; this ensures we do not over-fit the model (Neter, 1996:437). We check the model's robustness using the same validation data as for the logistic regression (Sipple, 2002: 72).

Utilizing the methodology found in Sipple's study, we build two regression models that we briefly introduce in this paragraph. We build one logistic model using 97 data points. This model predicts whether a program will have procurement cost growth. We then build a multiple regression model from the 53 of the 97 data points that have procurement cost growth. We apply a log transformation to the response variable in this model to correct for heteroskedasticity in the residual plot (Sipple, 2002: 72). We further explain the rationale and implications associated with this transformation in the next chapter.

Review of Methodology

This chapter explores the analytical procedures carried forth in this thesis. Herein, we establish the tie between the literature review and the analysis we perform. Additionally, we analyze the credibility of the SAR data, describe our data collection and compilation process, and describe the predictor variables that we investigate in our models. Finally, we explain the rationale for our use of logistic and multiple regression techniques and the process into which we incorporate these techniques (Sipple, 2002: 72).

IV. Results

Chapter Overview

This chapter reports the results of both the logistic and multiple regression analysis. First, we discuss the processes underlying model construction and selection. Then, we review the resulting models and analyze their robustness. Finally, we evaluate the models for validity and practical usefulness. We examine both models (logistic and multiple) for each number of predictor variables we use. We name the resulting models after the type of regression and number of variables. For example, L.1 refers to the logistic regression model that uses only one predictor variable, and M.5 refers to the multiple regression model that has five predictor variables using data from only those programs that have cost growth.

Preemptive Data Analysis

As discussed in Chapter III, we set forth in this research to develop a model that predicts procurement cost growth during the EMD phase of development. A preliminary analysis of the response variable *Procurement Cost Growth %* via Figure 3 indicates a mixed distribution; this distribution contains a discrete mass at zero (24 data points) and displays a continuous distribution elsewhere. These findings are identical to those established by Sipple (2002). Therefore, we employ the methodology established in that study. We develop a two-step model utilizing both multiple and logistic regression techniques to ensure statistically valid results. Thus, we formulate a logistic regression model for use in determining whether a program incurs cost growth or not. We follow with a multiple regression model developed to determine the amount of cost growth that occurs given that the logistic model has predicted cost growth.

Stem	Leaf	Count
13		
12	0	1
11		
10		
9		
8		
7		
6		
5	6	1
4	29	2
3	236	3
2	78	2
1	0001244555	10
0	1111111111222222333333334444455566666677889	43
-0	9987775543333322222111111000000000000000000000000	49
-1	00000000000	11

Figure 3. Stem-and-Leaf Plot of Y (Increments; Stem = 10%, Leaf = 1%)

Logistic Regression Results

The immense number of possible predictor variable combinations makes finding a true "best" model an unattainable goal, so we set out to produce the most predictive model possible within our resource constraints. Given the enormity of exploring all of the possible combinations, we narrow our predictor combinations to only those that show the most promise as we progress from a single-variable model to a three-variable model. We begin by regressing all one-variable models and recording the results. From these results, we select the ten "best" one-variable models and regress all possible two-variable models that stem from each of those models. Next, we select the nine models that appear most significant from the two-variable results and regress all possible three-variable models that stem from each of those models. To ensure that we do not violate the established data point to variable ratio, we refrain from building models that utilize more than three variables.

Each generation of regressions presents us with several candidate models to be carried forward for regression with additional variables. Within each of these generations, we then compare the candidate models and identify the best model. Table 5 summarizes the pertinent statistical characteristics of the "best" models. We select these models over other candidate models based on the measures listed in the table. The following paragraphs discuss these measures.

Table 5. Evaluation Measures for Logistic Regression

Evaluation Measures			
Number of Predictor Variables	1	2	3
R^2 (U)	0.2456	0.4975	0.8307
Number of Data Points	97	35	35
Area Under ROC	0.81517	0.91608	0.99301

First we compare models based on the uncertainty coefficient or R^2 (U). 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. This R^2 (U) statistic "is the proportion of the total uncertainty that is attributed to the model fit (JMP® 5.0, 2002: Help)". As with ordinary least squares (OLS) a value of 0 indicates a weak model and that the Xs have no predictive effect, while an R^2 (U) of 1 indicates a perfect fit. The models we select all have the highest R^2 (U)s of any of the other models within the same generation of predictors. It is important to note that although there are similarities, the interpretation of the R^2 (U) differs from the R^2 of linear models (OLS). For an in-depth explanation of R^2 (U) consult Sipple (2002).

Next, we consider the number of data points. The number of data points a model utilizes is particularly important for two reasons. First, the larger the sample size, the more of our population we capture in our sample. Second, the greater the number of data points, the more predictor variables we can add before the model becomes invalid statistically. According to Neter et al., a model should have at least six to ten data points for every predictor used (Neter, 1996:437). A significant decrease in the number of data points to predictor variables occurs when we incorporate *FUE-based Maturity %* into the two-variable model. We find this decrease in data points to be an acceptable tradeoff considering the increase in the models predictive capability when this variable is added. Further, the model still meets our established guidelines for the ratio of data points to predictor variables. In fact, all logistic models developed in this study have more than ten data points for every predictor used.

Third, we consider the area under the receiver operating characteristic (ROC) curve. According to the JMP® help menu, the ROC curve maps out the proportion of the true positives (sensitivity) out of all actual positives versus the proportion of false positives (1-specificity) out of actual negatives, both calculated across all possible calibrations of the model. We classify a true positive as a program incurring cost growth when the model predicts that cost growth will occur. Further, we define a false positive when the model predicts that cost growth will occur, but the program does not incur cost growth. The area under the ROC curve, then, gives an idea of the probability associated with ability of the model to accurately predict whether a program will have cost growth, based on results from the fitted values (Goodman, 1998:Appendix A).

This evaluation measure deals specifically with our goal of accurately assessing whether a program will or will not have cost growth. In reality, cost estimators rarely concern themselves with false positives, because predicting cost growth for a program that does not incur cost growth causes few problems. Armed with this knowledge, we seek to provide the most accurate model possible, and thus, we minimize all model errors when possible, including false positives.

We consider the overall predictive ability of each candidate model as our fourth statistical measure of interest. For this evaluation, we focus on the *p*-value associated with the Chi-squared test. JMP® 's on-line help provides the following interpretation: The Chi Square tests the null hypothesis that all regression parameters are zero (have no predictive ability), and that "it is computed by taking twice the difference in the negative log likelihoods between the fitted and reduced model that has only the intercepts." The resulting *p*-value that a model exhibits is "the probability of obtaining a greater Chi Square value by chance alone" (JMP®, 2002: Help). We find that all logistic regression models have a *p*-value less than 0.0001. Therefore, we cannot use this measure to further differentiate between models.

In addition to whole-model statistics, we consider the *p*-values for the parameter estimates. A lower *p*-value indicates higher statistical significance for that parameter as an estimator of the response variable. We desire the *p*-values to be less than 0.05 in order to ensure parameter significance. Additionally, we prefer the *p*-values to be as low as possible; this precautionary measure helps prevent tailoring the model to the fitted data to the extent that it lessens the ability of the model to predict the response values of the population (Sipple, 2002: 81). The two- and three-variable models in Table 6 breach the 0.05 criterion. *FUE-based Maturity* appears insignificant (0.1285) in L.2 and borderline significant (0.0594) in L.3. Although this variable breaches the established criteria, the model proves to be the most significant at both the two- and three-variable level. Another variable, *Class S,* appears borderline (0.0689) in L.3 as well. Since both variables are borderline significant in L.3, we do not disqualify these variables as a candidate estimators. Thus, we consider all the models listed in Table 5 as potential candidates for modeling whether a program will have cost growth.

Table 6. *P*-Values of Predictor Variables for Logistic Model

Number of Predictors			
Predictor Variables	1	2	3
Length of Production in Funding Yrs	0.0001	0.0053	0.0349
FUE-based Maturity		0.1285	0.0594
Class S			0.0689

To this point, our efforts focus on how the individual models fare against the established evaluation standards, but selecting a best model requires comparing between the models. Table 7 illustrates the combined impact that the incremental addition of predictors has on the various evaluation measures. This table shows the increase or decrease in each evaluation measure associated with the addition of a single predictor variable to the model. For example, as we add a predictor to L.1, we gain 0.2519 in R^2 (U), 0.10037 in area under ROC curve, and our ratio of data points to the number of independent variables in the model decreases to 17.

Table 7. Incremental Changes in Evaluation Measures for Logistic Model

Evaluation Measures			
Number of Predictors	1	2	3
R^2 (U)	0.2456	0.2519	0.3332
Data/Variable Ratio	97	17.5	11.6
Area Under ROC Curve	0.8152	0.10037	0.07693

Figure **4** visually depicts the effects on the whole-model statistics with each one-predictor increase. In this graph, an increasing trend line indicates that the addition of the extra predictor increases the predictive capability of the model. A flat, or decreasing trend line indicates that the addition of the extra predictor variable does not increase the predictive capability of the model. From the graph, we see similarities in the behavior of the whole-model measures. We establish that the addition of predictor variables significantly increases the predictive capabilities of the model through the addition of the third variable. For these reasons, we preliminarily consider L.3 as the best logistic model, based on the whole-model measures (Appendix A). Validation of the models will determine whether this initial conclusion endures.

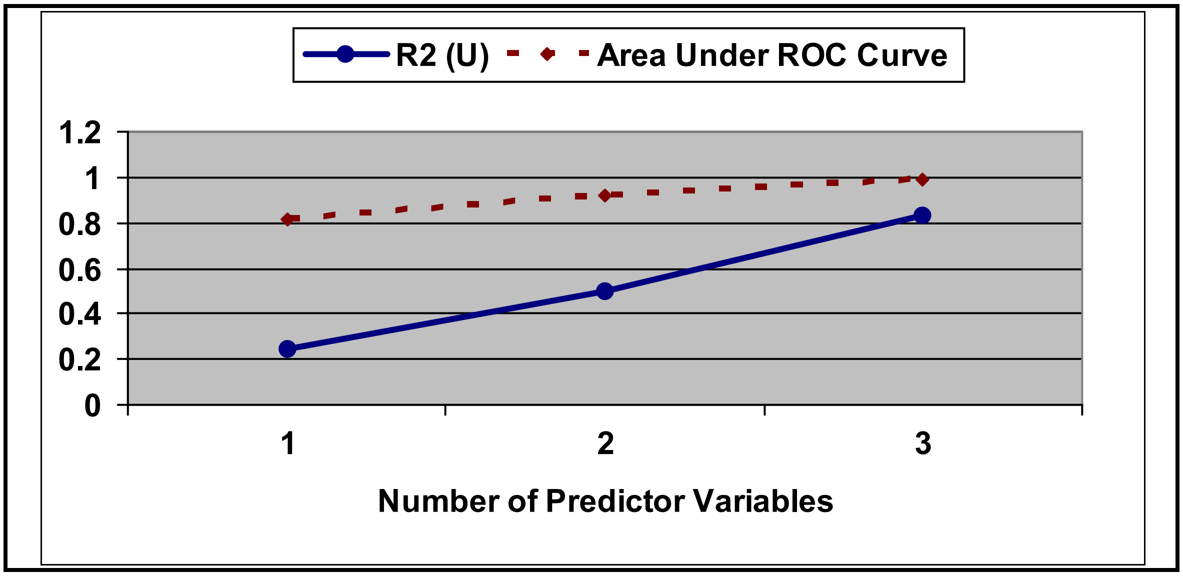


Figure 4. Incremental Changes for the Logistic Model (Cumulative)

It should be noted that we consider four-variable models, but the increase in predictor variables causes two problems. First, this increase in variables brings us below the 10 to 1 data point to variable ratio that we desire. This problem alone does not prevent us from considering four variable models. Second, we attempt to construct multiple four-variable models but these models exceed the capabilities provided by JMP®. Specifically, JMP® returns a "Failed to Converge" error message when additional variables are added.

For validation, we use 25 data points that we randomly select from the original 122-point data set. Of these 25 data points, 21 data points have missing values for some of the variables (namely FUE-based Maturity), leaving only 4 data points for validation. These 4 data points represent approximately 11 percent of the 39 useable data points the model incorporates. We initially establish a goal of validating 20 percent of the data, so we fall short of achieving the desired number. Therefore, we enter into the validation process concerned with the implications that accompany our limited validation pool, but defer further action until we complete the analysis.

The validation process entails saving the functionally predicted values (‘0’ or ‘1’) in JMP® for each of the validation data points and comparing those predicted values to the actual values. JMP® computes the predicted values by assessing the probability of having cost growth. JMP® assigns a ‘1’ to any point with a probability of 0.5 or greater and a ‘0’ otherwise (Sipple, 2002: 85). Upon validation, the model accurately predicts four out of the four data points for a success rate of 100 percent. We now consider the significance that the small number of validation data points imposes on our results and contemplate more extensive validation measures. Upon further analysis, we find the model to be accurate for 37 out of the 39 useable data points, further establishing that this model has some predictive ability, and confirming its place as our best model (Appendix A). From these results, we deem additional validation measures unnecessary. Table 8 displays the validation results for all 39 data points.

Table 8. Validation for Logistic Regression

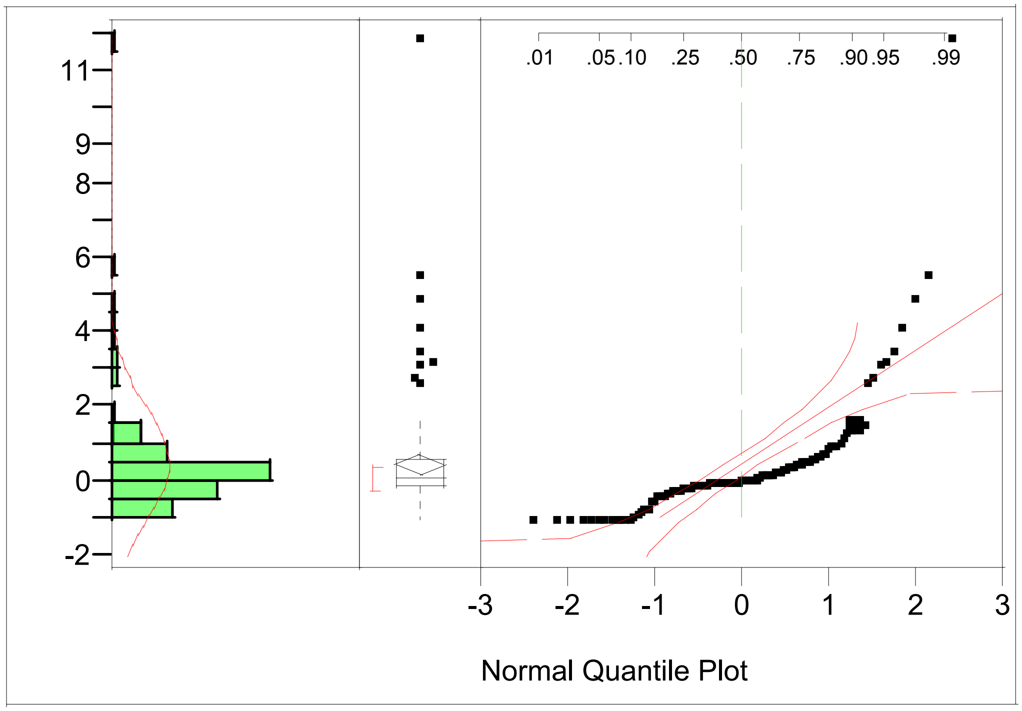
Program	Predicted	Actual
CGS (JSTARS GSM)	0	0
CSSCS (ARMY)	1	0
E-2C Computer Upgrade	0	0
E-6A TACAMO (NAVY-COMM)	0	0
FAAD C2I	0	0
FAAD NLOS Fiber Optic Guided-Missile	0	0
IAV	0	0
Javelin (AAWS-M)	0	0
JSIPS CIGSS	0	0
MLRS Upgrade Launcher	0	0
PLS (FHTV) (ARMY)	0	0
THAAD	0	0
Tomahawk TBIP	0	0
Uh-60 Upgrade (UH-60M)	0	0
ABRAMS Tank (M1,M1A1, & M1A2)	1	1
AFATDS	1	1
AH-64 Apache	1	1
Army TACMS (MGM-140A ATACMS)	1	1
BFVS A3 Upgrade	0	1
CH-47D Chinook	1	1
CH-47F (ICH)	1	1
FMTV	1	1
Harpoon A/R/UGM-84	1	1
JSOW BLU-108 (AGM-154B)	1	1
JSTARS (AIR FORCE)	1	1
Laser Hellfire	1	1
LHD-1	1	1
Longbow Apache Airframe Mods	1	1
Longbow Apache FCR	1	1
Longbow Hellfire	1	1
M1A2 Abrams Uprgrade	1	1
MMIII GRP	1	1
NAS	1	1
NAVSTAR User Equip	1	1
Navy Area TMBD	1	1
NSSN New Attack Sub	1	1
OH-58D Kiowa Warrior	1	1
Patriot PAC-3	1	1
Titan IV (CELV)	1	1

Multiple Regression Results

We build the multiple regression model for those occasions where a decision maker knows a program will have cost growth and wants to predict the amount of incurred cost growth. We begin model construction with our randomly selected 97 data points and exclude programs that have negative or no cost growth, leaving us with 55 data points. Focusing our efforts on only these points increases the models prediction accuracy, because it prevents data points outside the range of interest from skewing the results (Sipple, 2002: 86). We utilize the same 78-predictor variables as in logistic regression and we consider all possible interactions between variables. For the response variable (Y) we use *Procurement Cost Growth %,* which measures the percent increase of procurement cost growth from the Development Estimate.

We perform a preliminary analysis of the response variable to ensure that it is continuous in nature. From the results (Figure 4), we determine that the *Y*-variable exhibits a lognormal distribution. We perform a few test regressions and analyze the resulting residual plots (Figure 4). The plots fail to pass the visual inspection for constant variance as well as the Breusch-Pagan test (Neter, 1996: 112) at an alpha level of 0.05. Based on these findings, we transform the *Y* variable by taking the natural log. This transformation successfully removes the heteroskedasticity previously found and results in a distribution shape that is approximately normal (Figure 5). The distribution also passes the Shapiro-Wilk Test (JMP® 5.0, 2002: Help) for normality at an alpha level of 0.05.

Procurement Cost Growth %



Shapiro-Wilk W Test

W	Prob<W
0.628944	0.0000

Residual Plot

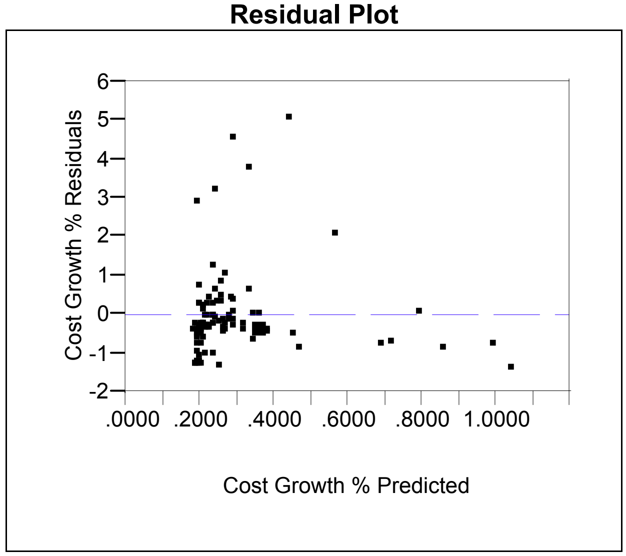
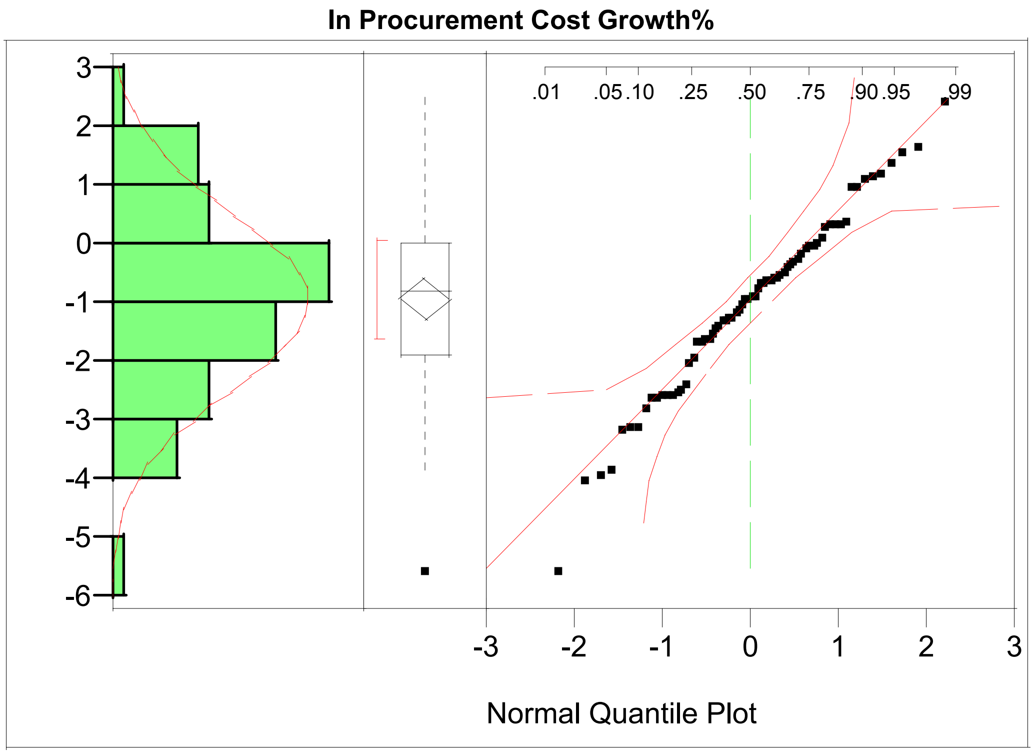


Figure 5. Distribution of *Y* and Residual Plot of Untransformed Model



Shapiro-Wilk W Test

W	Prob<W
0.985213	0.8503

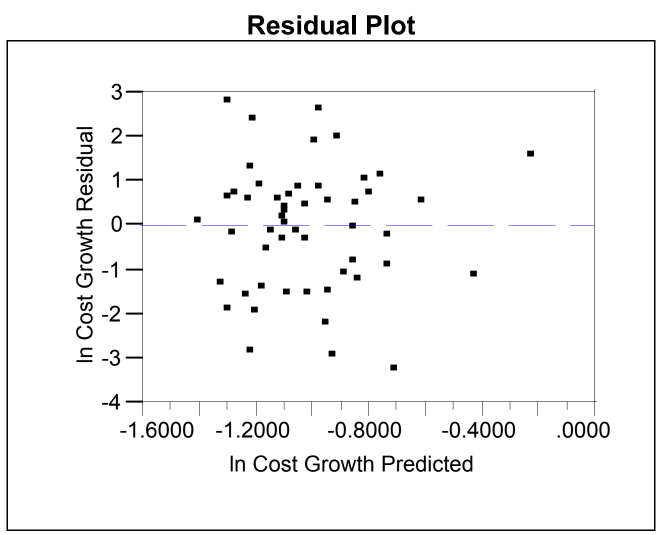


Figure 6. Distribution of Log *Y* and Residual Plot of Transformed Model

We utilize the automated stepwise regression found in JMP® to aid us in narrowing the number of possible predictor variable combinations. Since we start with only 55 data points, we limit the number of predictors to six in order to prevent the predictor to data point ratio from going too far below ten to one (Neter et al., 1996:437). Additionally, since we consider all variable interactions, we further constrain all models to contain at least three variables. We then analyze a multitude of regression models for each number of predictors (from three to six), just as we do for logistic regression. Finally, we choose the model that appears to provide the best prediction capability without violating any underlying statistical assumptions. Table 9 summarizes the best models for each generation of variables.

Table 9. Evaluation Measures for Multiple Regression

Evaluation Measures				
Number of Predictors	3	4	5	6
Adj R^2	0.594562	0.450139	0.45216	0.522666
Number of Data Points	22	51	51	51

We analyze the models to ensure compliance with the underlying assumptions of constant variance, normality, and independence. We find all models meet normality and constant variance assumptions at an alpha level of 0.05. Further, we removed all dependent programs during our initial data scrubbing and we find no obvious serial correlation present. Consequently, we assume independence within the data set. As an additional precaution, we test all predictors for multicollinearity by ensuring all variance inflation factors (VIFs) as calculated by JMP® are less than ten (Neter, 1996:387).

Model selection is based on the optimal mix of the statistical measures listed in Table 9. The evaluation measures for multiple regression are similar to those for logistic regression except we focus on Adjusted R^2 instead of R^2 (U). We choose the adjusted R^2 to measure the model's predictive ability over the standard R^2 because of its conservative nature. The R^2 value is subject to artificial inflation from simply adding additional variables to the model. Adjusted R^2 penalizes the model builder for adding variables that do not significantly increase the models predictive ability. Thus, by utilizing Adjusted R^2, we ensure that the variables within our model are significant.

From Table 9, we see certain patterns. First, as the number of predictor variables increases from three to four, the adjusted R^2 decreases but it increases thereafter. Also, the number of viable data points drops to 22 at the three-variable model, but it returns to 51 thereafter. Further, the adjusted R^2 decreases as we progress from M.3 to M.4, but increases from M.4 to M.6. The fluctuations that occur when moving from M.3 to M.4 are directly related to one variable. Model M.3 incorporates the variable *FUE-based Length of EMD*, which greatly increases the adjusted R^2, but drastically reduces the number of useable data points.

To remain near our initial goal of ten data points to each variable, we do not include this variable in any other models. The analysis of variance (ANOVA) *p*-value remains constant for all generations of predictors and therefore is not a discriminating factor in model comparison. The significance levels of the individual predictor variables are influenced by the interactions used in the models, but all non-interaction predictors significantly add to the model at an alpha level of 0.05. As with logistic regression, we chart the changes in the whole model evaluation measures (Table 10).

Table 10. Incremental Changes in Multiple Regression Models

Evaluation Measures				
Number of Predictors	3	4	5	6
Incremental increase in Adj R^2	0.594562	-0.144423	0.002021	0.070506
Ratio of data points to number of variables	7.3	12.75	10.2	8.5

From Table 10, we notice the largest marginal increase in adjusted R^2 occurs at the M.3 and the smallest at M.5. A fourth variable decreases adjusted R^2 by 0.144423. This decrease in adjusted R^2 does not call for the addition of a fourth variable. Conversely, the addition of the fifth and sixth variables increases the adjusted R^2, and therefore warrants the addition. We initially determine that pursuing a model with more than six variables would violate the proposed data point to variable ratio. Thus, we now compare the two most predictive models. Both M.3 and M.6 violate the established guidelines for data points to variables, so we turn to the next measure of differentiation, adjusted R^2. Model M.3 produces a significantly higher adjusted R^2 than M.6, so we preliminarily consider Model M.3 as the best model. As with logistic regression, validation of these models will determine whether this conclusion holds true.

For validation, we use the same validation data as for logistic regression. Only 17 of the original 25 validation data points have cost growth; the other 8 do not. The 17 represent approximately 25 percent of the programs within the data that contain cost growth. Therefore, we feel reasonably confident in the validation results. During model validation, we find that M.3 only uses 4 of the 17 data points because of missing data for some of the predictor variables (specifically, FUE-based length of EMD). These results are not surprising as they mirror the results from logistic regression. Thus, we feel confident proceeding with the validation process. To further ensure the validity of the results, we perform validation on 100 percent of the data set just as with the logistic regression model.

We create an upper bound for validation as opposed to a prediction interval for practicality reasons. In the cost-estimating environment very few decision makers are concerned with having too much money. Consequently, our goal is accurately predict the amount of cost growth while ensuring that the program is not underestimated. We consider an 80 percent upper prediction bounds. For an 80 percent upper bound, we expect to see approximately 80 percent of the validation data points fall under the bound. From the results of our validation, (Table 11) we determine that for the validation data our model is 100 percent accurate at a confidence bound of 80 percent. We are reasonably confident with these results. Thus, we find that Model M.3 most accurately predicts cost growth. (see Appendix B for model).

Table 11. Validation for Model M.3

Program	Upper Bound	Proc. Cost Growth %	Under Bound(=1)
AFATDS	0.29823463	0.02044542	1
BFVS A3 Upgrade	1.06506215	0.06539182	1
NSSN New Attack Sub	0.34067406	0.07603231	1
JSTARS (AIR FORCE)	0.70798088	0.13743423	1
Longbow Apache Airframe Mods	1.06506215	0.19645043	1
NAVSTAR User Equip	0.68274693	0.23135577	1
Longbow Hellfire	1.06506215	0.25796573	1
M1A2 Abrams Uprgrade	0.98674861	0.32678387	1
OH-58D Kiowa Warrior	0.43657577	0.34797855	1
Longbow Apache FCR	0.46882195	0.38306452	1
FMTV	1.62227683	0.40948964	1
Navy Area TMBD	1.5900236	0.43547886	1
Army TACMS (MGM-140A ATACMS)	0.63921818	0.50230742	1
NAS	1.37714478	0.5389487	1
MMIII GRP	3.21460278	0.56099202	1
CH-47D Chinook	1.12873534	0.63318452	1
JSOW BLU-108 (AGM-154B)	2.46525438	0.96972065	1
Patriot PAC-3	1.77084257	1.0265881	1
CH-47F (ICH)	1.46315307	1.19511582	1
Harpoon A/R/UGM-84	8.07029151	1.38891013	1
AH-64 Apache	1.96291705	1.44902572	1
LHD-1	2.20498034	1.48798368	1
Laser Hellfire	2.23637851	1.54969281	1
ABRAMS Tank (M1,M1A1, & M1A2)	14.9345382	2.73540905	1
Titan IV (CELV)	18.1478484	5.56894576	1

Chapter Summary

We analyze both logistic and multiple regression models in this chapter, each with several generations of sub models that differ in the number of variables used and the particular variables used. From these subsets we select the best models for each number of predictor variables and compare them using statistical measures of accuracy and significance until we arrive at a single best model for each family (Sipple, 2002: 125). We judge Models L.3 and M.3 as the best models for each family of model. Our study determines that these models perform reasonably well in determining whether a program will have cost growth and how much cost growth a program will have, respectively.

V. Discussion and Conclusions

Chapter Overview

This chapter summarizes the issues concerning cost growth in DoD acquisitions and the research efforts carried forth herein. First, we provide an overview of the problems facing the acquisition and cost estimating environments. Next, we summarize the results from the literature review. We then briefly run through the methodology employed during this study. We follow with a restatement of the achieved results, which we accompany with a list of practical limitations. Lastly, we provide recommendations for the implementation of this research as well as some possible areas for further research.

Explanation of the Issues

Cost growth continues to plague major acquisition programs in DoD. DoD's current reform focuses on improving the accuracy of the cost estimate. These efforts require that cost estimators provide more precise estimates by incorporating cost risk factors in their initial estimates to accommodate expected cost growth. Cost estimators currently have two sources for estimating cost risk factors, "expert opinion" and "historical data." Most cost analysts agree that the best sources for cost estimates come from relationships developed from recent, relevant, and accurate, historical databases. Hence, we logically conclude that the best sources for cost-growth estimates would also involve relationships developed from historical databases. In an effort to construct more accurate estimates, we pursue such relationships from the SAR database by utilizing both logistic and multiple regression techniques.

Summary of Literature Review Results

An extensive review of historical cost-growth studies in major defense acquisitions supports the research carried forth in this thesis. We reference a multitude of previous studies, but most vastly differ in scope from our study. Thus, we restrict our focus to a single study, Sipple 2002, as it proves most relevant to our research. Sipple's research provides us with a database, a methodology, and a list of 78 candidate predictor variables from which we springboard. The Sipple study focuses on the EMD phase of acquisition. Sipple further scopes his analysis to contain a single, SAR-defined category of cost growth, specifically, *Engineering* cost growth. The scope of this study differs from Sipple's, in that we focus on total procurement cost growth during the EMD phase of development, and do not analyze any individual SAR categories of cost growth. Although the differences between this study and its predecessor are slight, we still consider the applicability of the results with an appropriate degree of discretion.

Review of Methodologies

We use SARs as the sole source of information for purposes of analysis. We use the most current SARs to update the research database constructed by Sipple (2002). The most recent SARs available are from December 2001. Thus, the updated database now spans from 1990 through 2001. Additionally, to ensure independence of data points, we only include the most recent SAR for each program. Further, to avoid the confounding effects of inflation, we convert all dollar amounts into base year 2001 dollars. We then compute our response variable, which we call *Procurement Cost Growth %* for all programs. This variable represents the total cost variance in procurement dollars divided by the total baseline cost of a program in procurement dollars.

Once we review and update the database, we begin our preliminary analysis. Based on Sipple's findings (Sipple, 2002: 59), we expect the response variable to have a mixed distribution: about half of the data is massed at zero, while the other half is from a continuous distribution. An initial analysis of our data shows that the distribution for procurement cost growth during EMD is, in fact, from a mixed distribution. We therefore duplicate the procedures Sipple established in his research.

We first split the data into discrete and continuous distributions. We follow by utilizing logistic regression to analyze the discrete distribution and to discriminate between those programs that show cost growth and those that do not (we group negative cost variance with the latter). Once we determine that a program experiences cost growth, we utilize multiple regression to determine the amount of incurred cost growth.

As we begin to construct the multiple regression models, we find that *Procurement Cost Growth %* is from a lognormal distribution. We perform some test regressions and analyze the resulting residual plots. The plots fail to pass the visual inspection for constant variance as well as the Breusch-Pagan test at an alpha level of 0.05. Based on these findings, we transform the response variable via the natural log. This transformation successfully removes the heteroskedasticity previously found and results in a distribution that is approximately normal. Finally, to ensure that we do not overfit the models to the data, we set aside approximately 20 percent of the data for validation. Thus, we use the remaining 80 percent for model building.

Restatement of Results

We find that a three-variable model produces the best results for the logistic regression. This model accurately predicts 100 percent of the validation data and approximately 95 percent of the total data. Additionally, we conclude that the three-variable model produces the best results for multiple regression. At an 80 percent upper confidence bound, the model predicts correctly for 100 percent of the data. A correct prediction for this model infers that the actual amount of cost growth incurred is less than the predicted upper bound. Both the logistic and the multiple regression models satisfy all underlying statistical assumptions.

Our results not only validate the two-step methodology established by Sipple (2002), but they also provide insight into program characteristics that can be useful to predict procurement cost growth. Overall, FUE-based variables prove to be most significant and appear to greatly influence the predictive nature of the models. The FUE-based maturity of a program appears to be a strong indicator of whether a program will incur procurement cost growth. This relationship is intuitive because the further along a program is, the more likely the program is to have incurred cost growth. Additionally, the FUE-based length of EMD significantly influences how much cost growth a program will incur. This relationship is logical as well, since the longer the length of EMD, the more opportunities that there are for cost growth. We would expect all schedule variables to produce similar results, but FUE-based variables repeatedly produce superior results throughout our research efforts. These relationships should be further investigated in future research efforts. By investigating these predictive relationships, we add contemporary insight into the underlying drivers of procurement cost growth.

Limitations

Program cost growth at the aggregate level has proven difficult, if not impossible, to predict. Additionally, constructing a model to predict cost growth for a single SAR-defined category is of little use to the cost estimating community. Thus, we divide cost growth into the largest logical segments. For our research efforts, we address cost growth in procurement dollars and only in the EMD phase of acquisition. The resulting Cost Estimating Relationships only apply within the range of data used to construct them. Therefore, any use beyond these bounds may produce errant end results. Finally, the FUE-based variables were not available for a majority of programs, further limiting the applicability of these results.

Recommendations

The results from this study further validate the use of logistic regression in cost estimation. Sipple provides the following rationale for the implementation of logistic regression in the cost estimating community: First, logistic regression offers the ability to predict whether or not a program will experience cost growth. Second, logistic regression also provides an estimated probability that the program will have cost growth. Finally, logistic regression alleviates the estimator from attempting to interpret negative cost-growth results (Sipple, 2002:119). We further recommend the use of logistic regression in cost estimation. Multiple regression requires that the response variable be from a continuous distribution and cost data in general appears to originate from a mixture distribution. Therefore, we reason that logistic regression is required to ensure the validity of the model's results.

In situations where an estimator knows procurement cost growth exists, the multiple regression model (M.3) not only satisfies statistical requirements, but also predicts reliable upper bounds. The cost estimating community should consider this model when estimating procurement cost growth during EMD. As a cautionary note, this model only has utility in estimating procurement cost growth in EMD. Thus other models are necessary to fully account for program cost growth.

Possible Follow-on Theses:

We further encourage the exploitation of the database created during Sipple's research for other research topics. We present a wide range of data in order to facilitate the development of the predictor variables explored in this research. This database may prove useful in other cost related research and possibly even other programmatic areas. We provide the following possible examples:

- 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).
- Accomplish what we did for the PDRR and procurement phases for both RDT&E and procurement dollars (Sipple, 2002:121).
- Accomplish what we did for RDT&E dollars.
- Compare what we did with analyzing each SAR-category of procurement cost growth and then rolling them up into one estimate.
- Expand our research to include more programs, which should remove the problems we encountered with validation.
- Experiment with the sensitivity of the models we create to varying inputs (Sipple, 2002:121).

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

Appendix A. Logistic Regression Model (Model L.3)

Whole Model

RSquare (U)	0.8307
Observations (or Sum Wgts)	35
Prob>ChiSq	<0.0001
Area Under ROC Curve	0.99301

Parameter Estimates

Term	Estimate	Prob>ChiSq
Intercept	21.6061043	0.0349
Class - S	-9.5302264	0.0689
Length of Prod in Funding Yrs	-1.0997951	0.0390
Maturity of EMD (Maturity from MSII / FUE-based length of EMD)	-8.5808767	0.0594

Appendix B. Y-Transformed Multiple Regression Model (M.3)

Whole Model

RSquare Adj	0.594562
Observations (or Sum Wgts)	22
F Ratio	7.1592
Prob > F	0.0011

Parameter Estimates

Term	Estimate	Prob>|t|
Intercept	-0.891569	0.0967
FUE-based Length of EMD	0.0013256	0.7787
Service = Army only	-0.098109	0.8030
FUE-based Length of EMD-92.6818)*(Service = Army only-0.54545)	0.0578596	0.0002
Electronic	-0.569309	0.2262
FUE-based Length of EMD -92.6818)*(Electronic-0.22727)	0.0321444	0.0189

Bibliography

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

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.

Coleman, R. L., J. R. Summerville, M. DuBois, and B. Myers. "Risk in Cost Estimating: General Introduction & The BMDO Approach." Briefing at the 33rd Annual DoD Cost Analysis Symposium. Williamsburg VA. 2 February 2000.

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).

Goodman, Clifford S. "TA101: Introduction to Health Care Technology Assessment." A Presentation Prepared for Health Care Professionals by The Lewin Group. Bethesda MD: US National Library of Medicine, http://www.nlm.nih.gov/nichsr/ta101/ta101_c1.htm. 30 June 2002.

Grossman, Elaine. "Defense Budgets Include $7.4 Billion Boost For Realistic Costing." Early Bird Article. Inside the Pentagon, February 7, 2002. https://www.asc.wpafb.af.mil/news/earlybird

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

JMP®. Version 4.0.4, IBM, 51.5M, disk. Computer software. SAS Institute Inc., Cary NC, 2000.

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

Knoche, Chris. "Defense Acquisition Deskbook: Selected Acquisition Report." Online Defense Acquisition Reference Guide. http://www.deskbook.osd.mil/default.asp. 30 June 2002.

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

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).

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

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

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
09-03-2003	**Master's Thesis**	Jun 2002 – Mar 2003

4. TITLE AND SUBTITLE	**5a. CONTRACT NUMBER**
ESTIMATING PROCUREMENT COST GROWTH USING LOGISTIC AND MULTIPLE REGRESSION	**5b. GRANT NUMBER**
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Moore, Gary, W., Captain, USAF	**5d. PROJECT NUMBER**
	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 P Street, Building 640 WPAFB OH 45433-7765	AFIT/GCA/ENC/03-02
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 WPAFB OH 45433-7123 DSN: 986-5478 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 effort validates, and further explores the use of a two-step procedure for assessing DoD major weapon system cost growth using historical data. We compile programmatic data from the Selected Acquisition Reports (SARs) between 1990 and 2001 for programs covering all defense departments. Our analysis concentrates on cost growth in the procurement dollar accounts for the Engineering and Manufacturing Development phase of acquisition. 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, a multiple regression step is implemented to predict how much cost growth will occur. Our study focuses on the seven SAR cost growth categories within the research and development accounts – economic, quantity, engineering, schedule, estimating, support, and other. We investigate these categories at the aggregate level for significant cost growth characteristics and develop predictive models where appropriate.

15. SUBJECT TERMS

Logistic Regression, Multiple Regression, Cost Variance, Cost Growth, Selected Acquisition Report, SAR, DoD Cost Growth, DoD Cost Variance Study, Inferential Statistics, Cost Growth in DoD Acquisition Programs, Predicting 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** Edward D. White, USAF, (ENC)
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	UU	97	(937) 255-3636; e-mail: Edward.White@AFIT.edu

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