9-1996

Calibration of the Checkpoint Model to the Space and Missile Systems Center (SMC) Software Database (SWDB)

Karen R. Mertes

CALIBRATION OF THE CHECKPOINT MODEL TO THE SPACE AND MISSILE SYSTEMS CENTER (SMC) SOFTWARE DATABASE (SWDB)

THESIS

Karen R. Mertes, Captain, USAF

AFIT/GCA/LAS/96S-11

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GCA/LAS/96S-11

CALIBRATION OF THE CHECKPOINT MODEL TO THE SPACE AND MISSILE SYSTEMS CENTER (SMC) SOFTWARE DATABASE (SWDB)

THESIS

Karen R. Mertes, Captain, USAF

AFIT/GCA/LAS/96S-11

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

AFIT/GCA/LAS/96S-11

CALIBRATION OF THE CHECKPOINT MODEL TO THE SPACE AND MISSILE SYSTEMS CENTER (SMC) SOFTWARE DATABASE (SWDB)

THESIS

Presented to the Faculty of the Graduate School of

Logistics and Acquisition Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Cost Analysis

Karen R. Mertes, B.A., M.S.B.A.

Captain, USAF

September 1996

Preface

This thesis effort covers the discovery, evaluation, and documentation process of the calibration procedure of the CHECKPOINT software cost estimation model. A calibration and validation of the model were performed using data points found in the Space and Missile Systems Center Software Database. The objective was to assess how well CHECKPOINT predicted the actual project development effort and schedule. The results of this effort are truly remarkable; they have broken the mold that software cost estimating models are only accurate to within 25% half of the time. In using CHECKPOINT, the software cost estimating arena now has a model whose reliability has surpassed all others. The results of this effort are highly significant and will positively impact the software community for months and possibly even years to come.

Acknowledgments

First and foremost, I wish to thank Professor Daniel Ferens for his guidance, wisdom, and enthusiasm regarding this project. I chose this topic because I have a tremendous amount of respect for my advisor. Both his knowledge and caring attitude helped turn my effort into a sound product. I would also like to thank Dr. David Christensen for devoting his time towards improving this product as my thesis reader. His insights helped make it flow!

Thank you to Mrs. Sherry Stukes and Ms. Shirley Tinkler for meeting with me during the process and reading the earlier versions of this product. Their helpful suggestions focused this effort in the direction that best benefited the software community.

Without the generosity of Mr. Capers Jones and Mr. John Zimmerman, both of SPR, Inc., this effort could not have been accomplished. I cannot thank SPR, Inc. enough for allowing me to use the model during the past nine months. In addition, without the visit by Mr. Zimmerman, the highly significant results may not have been obtained.

A special note of thanks are sent to Mr. Jim Bradley, Director, Depot Maintenance Systems Engineering, BDM Federal, Inc. for his continued support by providing insight and expertise throughout this endeavor. Our weekly meetings made the process less stressful!

A big thank you to my predecessors in "The Pentateuch Study" ... 1Lt Carl Vegas, Capt Robert Kressin, Capt Kolin Rathmann, Capt James Galonsky, and Mrs. Betty Weber. Thanks for setting the standard so high. It was a pleasure to follow in your footsteps and be part of the now, "Septuagint."

Last, but not least, my heartfelt thanks go to my husband and best friend, David Mertes. I hope you know I could not have completed this project without you. Your selflessness in allowing me to spend the needed time on this effort is appreciated more than words can say! Thanks so much for recovering my work when I lost sixty plus pages. Thanks to you I only lost a day of work instead of months! You're the greatest!

Karen R. Mertes

Table of Contents

Page

Preface ii

Acknowledgments iii

List of Figures xii

List of Tables xiii

List of Equations xiv

Abstract xv

I. Introduction 1

General Issue 1

Specific Issue 4

The Task 5

CHECKPOINT: The Model 6

Calibration 6

Importance 7

Research Objective 7

Scope of Research 8

Thesis Overview 9

Page

II. Literature Review 11

Overview 11

Cost Estimation Techniques 11

Description of Techniques 12

Technique Advantages and Disadvantages 12

Software Cost Estimation 14

Background 14

Uncertainty 15

Requirements Uncertainty 15

Cost Estimating Uncertainty 16

History of Function Points 17

Previous Calibration Efforts 20

The Thibodeau Study 20

The Illinois Institute of Technology Study 21

Ourada Thesis 22

The Pentateuch Study 22

Summary 24

III. Methodology 26

Overview 26

CHECKPOINT, Version 2.3.1 26

Data Description 27

Page

Stratification of SWDB 29

Assumptions 30

Normalization 30

Calibration 31

Calibration Procedure 32

Step-by-Step Calibration Procedure 32

Replication Instructions 32

Validation 34

Magnitude of Relative Error (MRE) 35

Mean Magnitude of Relative Error (MMRE) 35

Root Mean Square Error (RMS) 36

Relative Root Mean Square Error (RRMS) 37

Prediction at Level *k/n* (PRED(*l*) 37

Wilcoxon Signed-Rank Test 38

Summary 39

IV. Findings and Analysis 40

Overview 40

The Data 40

Identification 41

Page

The Results 44

MIS (Effort) 45

Calibrated Model 47

Uncalibrated Model 47

Comparison Between the Calibrated Model and the Uncalibrated Model 48

Military Mobile (Effort) 48

Calibrated Model 50

Uncalibrated Model 50

Comparison Between the Calibrated Model and the Uncalibrated Model 51

Military - Specific Avionics (Effort) 51

Calibrated Model 53

Uncalibrated Model 53

Comparison Between the Calibrated Model and the Uncalibrated Model 54

Military Ground and Application - Command and Control (Effort) 54

Calibrated Model 56

Uncalibrated Model 56

Comparison Between the Calibrated Model and the Uncalibrated Model 57

Page

Military Ground and Application - Signal Processing (Effort) 57

Calibrated Model 59

Uncalibrated Model 59

Comparison Between the Calibrated Model and the Uncalibrated Model 60

Unmanned Space (Effort) 60

Calibrated Model 62

Uncalibrated Model 62

Comparison Between the Calibrated Model and the Uncalibrated Model 63

Ground in Support of Space (Effort) 63

Calibrated Model 65

Uncalibrated Model 65

Comparison Between the Calibrated Model and the Uncalibrated Model 66

Cobol Projects (Effort) 66

Calibrated Model 68

Uncalibrated Model 68

Comparison Between the Calibrated Model and the Uncalibrated Model 69

Page

MIS (Schedule) 69

Calibrated Model 71

Uncalibrated Model 71

Comparison Between the Calibrated Model and the Uncalibrated Model 72

Unmanned Space (Schedule) 72

Calibrated Model 74

Uncalibrated Model 74

Comparison Between the Calibrated Model and the Uncalibrated Model 74

Ground in Support of Space (Schedule) 75

Calibrated Model 77

Uncalibrated Model 77

Comparison Between the Calibrated Model and the Uncalibrated Model 78

Cobol Projects (Schedule) 78

Calibrated Model 80

Uncalibrated Model 80

Comparison Between the Calibrated Model and the Uncalibrated Model 81

Synopsis of Wilcoxon Signed-Rank Test 81

Synopsis of Calibrated Function Point Categories 81

Synopsis of Calibrated Effort and SLOC Categories 82

Page

Synopsis of Calibrated Effort, SLOC, and Schedule Categories 83

Comparison of Calibrated Function Point Categories with Calibrated Effort and SLOC Categories 83

Comparison of Calibrated Function Point Categories with Calibrated Effort, SLOC and Schedule Categories 84

Synopsis of Calibrated Schedule Categories 84

Contrast Between Calibrated Effort and Calibrated Schedule 84

Analysis of the Calibrated Data Compared with the Uncalibrated Data ...84

Summary 86

V. Conclusions and Recommendations for Follow-on Research 87

Overview 87

Limitations 88

Summary of Results 88

Recommendations for Follow-on Research 90

Appendix A: Glossary 91

Appendix B: Data Records 94

Bibliography 116

Vita 119

List of Figures

Figure Page

1. MIS (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort 46

2. Military Mobile (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort 49

3. Military - Specific Avionics (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort 52

4. Military Ground and Application - Command and Control (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort 55

5. Military Ground and Application - Signal Processing (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort 58

6. Unmanned Space (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort 61

7. Ground in Support of Space (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort 64

8. Cobol Projects (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort 67

9. MIS (Schedule) Graph for Calibrated, Uncalibrated, and Actual Schedule.....70

10. Unmanned Space (Schedule) Graph for Calibrated, Uncalibrated, and Actual Schedule 73

11. Ground in Support of Space (Schedule) Graph for Calibrated, Uncalibrated, and Actual Schedule 76

12. Cobol Projects (Schedule) Graph for Calibrated, Uncalibrated, and Actual Schedule 79

List of Tables

Table		Page
1.	Strengths and Weaknesses of Software Estimation Methods	13
2.	Normalization of Effort and Schedule	31
3.	Summary of Calibration Categories, Methods, Languages, and Number of Data Points	43
4.	MIS (Effort)	46
5.	Military Mobile (Effort)	49
6.	Military - Specific Avionics (Effort)	52
7.	Military Ground and Application - Command and Control (Effort)	55
8.	Military Ground and Application - Signal Processing (Effort)	58
9.	Unmanned Space (Effort)	61
10.	Ground in Support of Space (Effort)	64
11.	Cobol Projects (Effort)	67
12.	MIS (Schedule)	70
13.	Unmanned Space (Schedule)	73
14.	Ground in Support of Space (Schedule)	76
15.	Cobol Projects (Schedule)	79
16.	Summary Statistics for the Eight Categories	89

List of Equations

Equation — Page

1. Magnitude of Relative Error 35
2. Mean Magnitude of Relative Error 36
3. Root Mean Square Error 36
4. Relative Root Mean Square Error 37
5. Prediction at Level *k/n* (PRED(*l*)) 37

Abstract

This study analyzed the effect of calibration on the performance of the CHECKPOINT Version 2.3.1 software cost estimating model. Data used for input into the model were drawn from the FY 95 USAF SMC Software Database (SWDB) Version 2.1. A comparison was made between the model's accuracy before and after calibration. This was done using records which were not used in calibration, referred to as validation points. A comparison of calibration points, both before and after, was done in order to assess whether calibration results in more consistency within the data set used. Six measures such as magnitude of relative error (MRE), mean magnitude of relative error (MMRE), root mean square error (RMS), relative root mean square error (RRMS), the prediction at level *k/n*, and the Wilcoxon Signed-Rank Test were used to describe accuracy. The results of this effort showed that calibration of the CHECKPOINT model can improve cost estimation accuracy for development effort by as much as 96.71%.

CALIBRATION OF THE CHECKPOINT MODEL TO THE SPACE AND MISSILE SYSTEMS CENTER (SMC) SOFTWARE DATABASE (SWDB)

I. Introduction

General Issue

Over the past fifty years, software has become a major factor in global business success, government operations, and in military strategy and tactics (Jones, 1994:99). The increasing usefulness of software in various applications is making the cost of software an increasingly greater portion of the total cost of computer systems (Boehm, 1984:17). Alone, the Department of Defense (DoD) currently spends $30 billion annually on developing and maintaining software (Ferens, 1995:2). The tremendous growth of our investment in software can be seen by comparing the DoD statistic to the fact that in 1980, the U.S. as a whole spent $40 billion on developing and maintaining software (Boehm, 1984:17). As a result, software has become a high visibility topic, and estimates of its development and maintenance costs in future projects are of concern to the DoD. While the maintenance or life-cycle costs of software are important, this research effort will limit its focus to development cost and schedule.

Software development is very labor intensive, requires long development schedules, and often suffers from the problem of distressingly low quality (Jones, 1994:99). Recent studies performed for the DoD environment by independent agencies (i.e., those not affiliated with any commercial cost model developer or marketer) have generally shown software cost estimation models to be, at best, accurate only to within 25% of actual cost or schedule about 50% of the time (Christensen/Ferens, 1995:1). With such lackluster results, senior company and government leaders have been dissatisfied with their investments in software cost estimation models. While it is recognized that software is a critical component of business and government operating efficiency, there is a widespread uncertainty about software cost estimation (Jones, 1994:99).

According to Bernard Londeix, an estimate of software development costs is successful when:

1) The early estimate is within +/- 30% of the actual final cost: this is the accuracy currently obtainable at an early stage of the development.

2) The method allows refinement of the estimate during the software life cycle. A higher accuracy can be achieved by monitoring and re-estimating the development each time more information is available.

3) The method is easy to use for an estimator. This enables a quick re-estimate whenever it is necessary; for example during a progress meeting, the evaluation of alternatives in strategic choices.

4) The rules are understood by everybody concerned. Management feels more secure when the estimating procedures are easily understandable.

5) The method is supported by tools and documented. The availability of tools increase the effectiveness of the method, mainly because results can be obtained more quickly and in a standard fashion.

6) The estimating process can be trusted by software development teams and their management. This helps in gaining the participation of everybody concerned with the estimate. (Londeix, 1987:3)

One of the problems inherent in software cost estimating models, identified by Jones, is its poor estimating accuracy. For this reason, more effort needs to be put into developing more accurate models and improving the performance of current models through calibration (which is defined in the "Glossary" in Appendix A). A procedure to improve the accuracy of a cost estimation model's ability to calculate software cost is to calibrate it to the user's environment. In fact, Thibodeau stated in his report, "We have shown that the calibration of model parameters may be as important as model structure in explaining estimating accuracy" (Thibodeau, 1981:5-29). In addition, his studies showed that calibration can improve accuracy by a factor of five (Thibodeau, 1981:5-29). Once calibrated, models can be accurate "to within 10%, 90% of the time" (Hayes, 1996).

According to Jones, the function point metrics are a concept for computing software size from five attributes: external inputs, external outputs, external inquiries, internal files, and external interfaces. Jones opines that function points are far superior to the source line of code metrics for expressing normalized productivity data. As real costs decline, cost per function point also declines. As real productivity goes up, function points per person month also go up. Function points would be the appropriate choice for a metric for applications where the number of algorithms is uncertain or where algorithmic factors are not significant. Accounting software, customer information systems, and marketing support systems are three examples of business applications that use function points (Ferens, 1995).

One model that supports the use of function points is CHECKPOINT. Like other cost estimating models, CHECKPOINT is comprised of algorithms which require the user to input values for certain parameters to calculate a specific software cost estimate. The parameters used in a given model vary and are those determined to be important by the creator of the model (in this case, Capers Jones of Software Productivity Research, Inc. (SPR)). Two other models, PRICE-S and SEER-SEM, can estimate software development costs using function points. However, these models first convert the function points to Source Lines of Code (SLOC) using standard conversion ratios based on language before estimating cost and schedule (Stukes, 1996). The essential differences among the various models is that each uses different cost estimating relationships in their algorithms and each emphasizes different parameters within those relationships (Vegas, 1995:3).

Specific Issue

This research effort adds an additional model to *The Pentateuch Study*. This study was initiated in August, 1994 as a sponsored AFIT thesis proposal from an Air Force Space and Missile Systems Center (SMC) organization, SMC/FMC, and Management Consulting and Research, Inc. (MCR). The objective of the study was to examine the calibration procedures and determine whether five selected models, PRICE-S, REVIC, SASET, SEER-SEM, and SLIM, could be calibrated to the SMC Software Database (SWDB). Improvements to estimating accuracy could be expected as a result of calibration. However, their results did not support this hypothesis. Five students in the

Graduate Cost Analysis program were tasked to calibrate the five models to the FY94 SMC SWDB as their thesis efforts (Christensen/Ferens, 1995:2).

The one significant difference among CHECKPOINT and three of the previously calibrated models is that CHECKPOINT can use either function points or SLOC (which are converted to function points via the ratio table in the CHECKPOINT User's Guide) as its inputs whereas the REVIC, SASET, and SLIM models can only have SLOC as their inputs. SEER-SEM and PRICE-S can also have function points as an input but these models were only calibrated using SLOC as part of *The Pentateuch Study*.

The Task. SMC is one of the many DoD organizations that invests heavily in computer systems. This is due to the high technology programs inherent in the space systems it manages. The precision of such systems leads to the need for highly reliable software. "High reliability" requires extensive testing and, consequently, is very expensive. SMC has, therefore, expressed an interest in having CHECKPOINT calibrated to the SWDB. This database contains historical development and to a limited extent, maintenance software data for previous projects which include:

- Military Mobile
- Military-Specific Avionics
- Military Ground and Application -- Command and Control
- Military Ground and Application -- Signal Processing
- Unmanned Space
- Ground in Support of Space and
- Missile

The primary purpose of this research effort is to aid DoD decision makers by providing a calibration method, based on the most current data available, that may improve CHECKPOINT's accuracy in predicting future project software effort (cost). Knowledge is power. Arming DoD decision makers with an accurate estimate of future software project costs will enable them to adequately plan the schedule for time sensitive and costly projects. It will also allow them to select from alternatives and feel confident in their decisions. The secondary goal of this effort is to provide the reader with a step-by-step reference of how to calibrate the current version of CHECKPOINT to their own database.

CHECKPOINT: The Model. The current version of CHECKPOINT is 2.3.1 which was released in February 1996. It was developed by Capers Jones of SPR. It is a PC-based software estimating cost model operating in a DOS environment available for purchase through SPR. A more detailed description is provided in Chapter III.

Calibration. Calibration is the adjustment of a model's equations to induce the model to provide a predicted outcome as close as possible to the actual outcome for a given set of data. CHECKPOINT will be calibrated using projects, profiles, and templates. Each data point will be entered as a project and the data points randomly identified for calibration in each of the eight categories will be saved to a profile. Eight categories will exist for this research effort, yielding eight profiles. Each profile will then be saved as a template (Zimmerman, 1996). Each category will be validated for accuracy using the following statistics: Magnitude of Relative Error (MRE), Mean Magnitude of

Relative Error (MMRE), Root Mean Square Error (RMS), Relative Root Mean Square Error (RRMS), Prediction at Level *k/n,* and the Wilcoxon Signed-Rank Test.

Importance

This research effort is ground breaking in that there have been no prior DoD attempts at calibrating CHECKPOINT to a database and publishing the results. The model has been calibrated by only a few corporations but they have not made their results available to the software community (Zimmerman, 1996). In addition, this effort was done using function points in three of the categories while effort, size, and schedule data were utilized in the remaining five categories. The fact that calibrations were performed on effort for all eight categories and on schedule for four of the categories is also unique. Calibrations on schedule have not yet been attempted and add a new dimension to *The Pentateuch Study* for further exploration. If the results prove significant, they will have a significant positive impact of the software community for years to come. The CHECKPOINT software cost estimating model may become the model of choice by even more user's in the field. Millions of dollars can be saved if the results prove significant.

Research Objective

In order to effectively calibrate CHECKPOINT, the following basic questions must be addressed:

1. What is CHECKPOINT'S pre-calibration accuracy with the data set selected for validation?
2. How is CHECKPOINT calibrated?

3. After calibrating CHECKPOINT with the selected data set, what is the model's accuracy with the validation data set?
4. What is the improvement in accuracy after calibration?

The above questions will be addressed by:

1. Learning and becoming knowledgeable about the CHECKPOINT cost estimating model, specifically with respect to calibration.
2. Obtaining and becoming familiar with the 1995 edition, version 2.1, of the SMC SWDB.
3. Stratifying the SMC SWDB into data sets to be calibrated by reviewing the data and organizing the SWDB into homogeneous groups of data.
4. Calibrating the CHECKPOINT model to each of the stratified data sets.
5. Determining the accuracy of and validating the calibrated model using the data reserved for this purpose.

Observations and conclusions will be made as well as recommendations regarding future calibration efforts for CHECKPOINT and the content of the SWDB as it pertains to CHECKPOINT calibration.

Scope of Research

The scope of this research effort is limited to calibration parameters derived for the operational environment reflected by the SMC SWDB described below. The results are SWDB specific; they represent only industry results that should not be automatically applied to other areas such as specific programs or contractors without conducting further calibrations unique to those programs. The SWDB is assumed to be reliable; inputs made

to subjective fields may contain rounding errors but these should be consistent throughout the 2,638 records contained in the SWDB. In addition, no inferences will be made as to the ability to calibrate any other model with this database since CHECKPOINT is not built on the same database as other models. The analysis accomplished in this research effort will encompass software development effort and schedule.

Thesis Overview

This research effort uses the SMC SWDB to calibrate the CHECKPOINT model. The results of the calibrated model are then validated.

Chapter II, Literature Review, reviews research efforts and literature in the area of software cost estimation. Specifically:

1) Software cost estimation techniques are reviewed. The advantages and disadvantages of each technique are then outlined.
2) A general background in the area of software cost estimation is given. The concept of uncertainty and its relationship to software cost estimation is presented. The use of function points as inputs to software cost estimation models is then described.
3) Previous calibration efforts are examined and their results presented.

Chapter III, Methodology, gives a general background and description of the CHECKPOINT model and the SMC SWDB, presents how the SMC SWDB was stratified, and shows how the CHECKPOINT model was calibrated with respect to effort and schedule. Additionally, this chapter outlines the specific methods used in validating the results and assessing the accuracy of the calibrated version of CHECKPOINT.

Chapter IV, Findings and Analysis, presents the results of the calibration effort with respect to the data and the model. This section presents the validation and statistical results.

Chapter V, Conclusions and Recommendations for Follow-on Research, provides observations and conclusions based on the findings and analysis discussed in Chapter IV. Recommendations in the area of calibration and possible follow-on research are provided.

II. Literature Review

Overview

Computers have touched private business and government alike. With the rising costs to develop software, the field of software cost estimating is rapidly gaining increased attention and respect among top leaders in organizations. This chapter reviews research efforts and literature in the area of software cost estimating. Specifically, this chapter addresses various cost estimation techniques, highlights the basic concepts and background of software cost estimation (including the area of uncertainty), describes a brief history of function points, and summarizes past calibration efforts.

Cost Estimation Techniques

There are several cost estimating techniques to choose from when estimating software costs. Generally, organizations will base their software cost estimates on past performance represented by historical data from which relevant cost factors are identified (Fairley, 1985:72). Historical data, however, are not always available or easily obtained. Typically, the cost estimation technique used depends upon the estimating organization's objectives, resources, and the basic capabilities and limitations associated with each technique.

Cost estimates are either top-down or bottom-up (Fairley, 1985:72). Top-down estimates emphasize total system costs first, whereas bottom-up estimates begin with the costs of lower sub-levels of the system and then aggregate these costs to derive a cost for the total system. Regardless of their basic approach, there are several methods available to cost analysts from which to derive the estimate.

Description of Techniques. Whether using the top-down or bottom-up approach, the major cost estimation methods used today are algorithmic models, expert judgment, analogy, Parkinson principle, and price-to-win (Boehm, 1984:7). The following is a brief description of each method:

1) Algorithmic models: "These methods provide one or more algorithms which produce a software cost estimate as a function of a number of variables which are considered to be the major cost drivers." (Boehm, 1984:7)

2) Expert judgment: This method is based on the experience of one or more experts, and relies on their opinion. (Boehm, 1984:7)

3) Analogy: This method is based on real projects. It relates the costs of similar past projects to estimate a new project. (Boehm, 1984:7)

4) Parkinson principle: "A Parkinson principle ('work expands to fill the available volume') is invoked to equate the cost estimate to the available resources." (Boehm, 1984:7)

5) Price-to-win: The cost estimate is based on the objective of winning the contract or job. (Boehm, 1984:7)

According to Boehm, (1981:337) and Ferens (1995:2), the Parkinson principle and price-to-win method do not produce reasonable estimates and are unacceptable. In addition, note that the five techniques listed above are usually applied in conjunction with either a top-down or a bottom-up approach toward estimating.

Technique Advantages and Disadvantages. Every cost estimation technique has its advantages and disadvantages, and the particular technique used depends largely on various limiting factors such as the availability of historical data, the objective of the estimate, available resources to derive the estimate, and the deadline for the estimate. Therefore, when using a particular cost estimation technique, the capabilities and

limitations of that technique should be considered. Table 1 outlines the basic strengths and weaknesses of the common cost estimation techniques. The top-down and bottom-up estimation methods are listed first followed by the five cost estimation techniques:

Table 1
Strengths and Weaknesses of Software Estimation Methods (Boehm, 1981:342)

Method	Strengths	Weaknesses
Top-down	**- system level focus** **- efficient**	**- less detailed basis** **- less stable**
Bottom-up	**- more detailed basis** **- more stable** **- fosters individual commitment**	**- may overlook system level costs** **- requires more effort**
Algorithmic Model	**- objective, repeatable, analyzable formula** **- efficient, good for sensitivity analysis** **- objectively calibrated to experience**	**- subjective inputs** **- assessment of exceptional circumstances** **- calibrated to past, not future**
Expert Judgment	**- assessment of representativeness, interactions, exceptional circumstances**	**- no better than participants** **- biases, incomplete recall**
Analogy	**- based on representative experience**	**- representativeness of experience**
Parkinson	**- correlates with some experience**	**- reinforces poor practice**
Price-to-win	**- often gets the contract**	**- generally produces large overruns**

Boehm suggests that no one method is better than the next in all aspects (Boehm, 1984:7). Given Boehm's recommendation and the variety of software cost estimation approaches as well as the individual strengths and weaknesses of each, more than one

estimation technique should probably be used, if possible, and the results examined to derive a more comprehensive estimate (Fairley, 1985:84).

Software Cost Estimation

Background. Software costs are rising at an enormous rate. Therefore, software development is continuously troubled with cost overruns and schedule slippages. Software cost estimations serve as measurements in the planning and controlling of software development (DeMarco, 1982:40). They assist management in making informed decisions before, during, and after the software development process. However, the software development process is not clear. In fact, achieving software excellence has been compared to training for the Olympics. Both require good preparation and daily practice (Jones, 1994:99).

This complex software development process involves several interrelated tasks (Londeix, 1987:1). As a result, numerous factors can affect the cost of software development. Several software cost models have been developed and are in use today to manage the complex process of software development and the factors that affect its cost. Before escalating software costs became a significant issue in software development, software cost estimates were derived from a percentage of the hardware they supported (Wellman, 1992:30). Unfortunately, the time when software costs represented less than forty percent of the total cost of software development is long gone as software costs currently account for about ninety percent of the total cost to the end user over its life cycle (Wellman, 1992:30).

Cost models today attempt to capture and quantify software development complexity through empirically derived mathematical functions that relate the estimate to several key cost related inputs. To further understand and appreciate the complexity associated with software development and cost estimating, the area of "uncertainty" needs to be addressed.

Uncertainty. Decision makers must constantly decide on the proper allocation of resources and evaluate alternative ways of doing business so as to generate profit for their firm. However, the majority of these decisions are made under conditions of uncertainty. This term refers to the degree decision makers are unsure of the impact of alternative choices presented to them and the resources required to undertake a particular alternative. It represents the difference between what is estimated and what actually happens. This definition is standard for variance which is the difference between planned and actual cost. Cost estimating assists the decision making process by attempting to account for and describe or quantify as much uncertainty as possible regarding alternative courses of action. The goal of cost estimating is to provide decision makers with necessary and explicit information so that informed decisions can be made. Two major sources of uncertainty exist in the field of cost estimating: requirements uncertainty and cost estimating uncertainty (Fisher, 1962:4).

Requirements Uncertainty. "Requirements uncertainty refers to variations in cost estimates stemming from changes in the configuration of the system being costed, and is the major source of uncertainty with respect to military systems" (Fisher, 1962:4). There is more uncertainty in a software cost estimate if the requirements are not well defined.

The idea of requirements uncertainty is best explained by Boehm. He writes that the accuracy of a cost estimate is increased as the software development effort progresses and requirements become more refined (Boehm, 1984:8). Naturally, one would expect to be more confident in an estimate of something required a month from now versus a year from now. The more uncertain the requirements, the more likely an estimate will be based on algorithms derived from historical data. No matter how much requirements uncertainty exists, there will also be a certain degree of cost estimating uncertainty present throughout the software development process.

Cost Estimating Uncertainty. "Cost estimating uncertainty refers to variations in cost estimates of a system or force where the configuration of the system or force is essentially constant" (Fisher, 1962:4). Cost estimating uncertainty generally arises due to cost analyst differences, errors in data, and errors in developing the cost estimating relationships expressed in the model (Fisher, 1962:4). Cost estimating uncertainty tends to be minimal as compared to requirements uncertainty; however, significant variability can occur in estimates as a result of the input that goes into the estimate (Fisher, 1962:6). The estimating dilemma, better known as "garbage-in, garbage-out estimating," states that an estimate is only as good as its input (DeMarco, 1982:17).

Software development is inundated with complex tasks and uncertainty, largely in the earlier stages of the development process. As such, software cost estimating is a process that takes place throughout the software development life cycle and provides a measurement and foundation on which management can base their planning and controlling decisions. Today's software cost models attempt to quantify and express as

much complexity and uncertainty as possible by developing and using algorithmic models based on historical data. But how much past projects reflect future ones becomes an issue by itself when deriving an estimate from an empirically based algorithm. Considering the advancement of technology in the area of software development and that the software development process is still maturing, deriving an estimate based on past projects could be misleading. As a result, the data used to construct the model should be updated periodically and the model calibrated to specific environments. In order to derive more accurate and useful estimates, better input into the estimating process needs to be used.

What exactly this input should be is the problem faced by cost analysts and software engineers (Fisher, 1962:17). The input, whether in SLOC or function points, and how the input is treated and defined by the model is what differentiates cost models.

History of Function Points

The standard economic definition of productivity is goods or services produced per unit of labor and expense. Until 1979, when Albrecht published his Function Point metric, there was never a software definition of exactly what "function point goods or services" were in the output of a software project (Jones, 1994:99).

A metric frequently used to determine software efficiency is "cost per line of source code," which unfortunately did not correlate to the economic definition of productivity. Software involves a substantial percentage of fixed costs that are not associated with coding. When more powerful programming languages are used, the result is to reduce the number of "units" that must be produced for a given program or system (Jones, 1994:99).

In the late 1970's, Albrecht suggested that the economic output unit of software projects should be valid for all languages, and should represent topics of concern to the users of the software. In short, he wished to measure the functionality of software. Albrecht believed that the visible external aspects of software that could be counted accurately consisted of five items: the inputs to the application, the outputs from it, inquiries by users, the data files that would be updated by the application, and the interfaces to other applications. After trial and error, empirical weighting factors were developed for the five items. The number of external inputs (EI) was weighted by 4, external outputs (EO) by 5, external inquires (EQ) by 4, internal data file updates (ILF) by 10, and external interfaces (EIF) by 7. The basic function point equation is 4 EI + 5 EO + 4 EQ + 10 ILF + 7 EIF equals the number of basic function points (Ferens, 1995). These weights represent the approximate difficulty of implementing each of the five factors (Jones, 1994:99).

In 1986, the non-profit International Function Point Users Group (IFPUG) was formed to assist in transmitting data and information about function points (Jones, 1994:99). Function points give software engineering researchers a way of sizing software through the analysis of the implemented functionality of a system from the user's point of view. Function points also provide a way to predict the number of source code statements that must be written for a program or system (Jones, 1994:99).

As previously noted, SLOC lack a standard definition for any major programming language, and there are more than 400 programming languages in use (Jones, 1994:99) but only about ten in common use by DoD agencies (Stukes, 1996). The software

literature and even the lines of code counting standards are equally divided between those using physical lines and those using logical statements as the basis for the SLOC metric (Jones, 1994:99). This metric is particularly dangerous, if used carelessly, because "the models that are not sensitive to language believe any efficiencies of the language are reflected in the complexity attributes" (Stukes, 1996).

The function point metric, an alternative to the SLOC metric, is growing in use and popularity for software cost estimation. Function points originated with the work of Albrecht as a methodology for estimating the size of a program by the number of functions the software was performing (Ferens and Gurner, 1994:49). Based on his research, Albrecht further hypothesized that function points may be an alternative to using SLOC to estimate the cost or effort required for software development (Ferens and Gurner, 1994:49).

The advantage of using function points is the total number of these points for an application does not change with the programming language (Jones, 1994:99). Now, it is possible to see the economic advantages of higher order languages such as Ada. Another advantage to the use of the function point metric is the continual improvement of function point theory and its practice by the IFPUG (Ferens and Gurner, 1994:49).

Unfortunately, most of these advantages are also accompanied by disadvantages. One problem with function points is that Albrecht's five attributes are sometimes hard to define and count (Ferens, 1995). Another disadvantage is that function points are not readily adaptable to the real-time or scientific environments because of their difficulty in being counted (Ferens, 1995). However, this disadvantage can be countered by using

feature points in which a sixth attribute, the number of unique algorithms (AL) is added to the basic equation as described earlier. The new equation becomes: 3AL + 4EI + 5EO + 4EQ + 7ILF + 7EIF.

Previous Calibration Efforts

Since calibration improves a software cost estimation model's accuracy in predicting development effort, numerous research has been done in the area of software cost model calibration. Though the results of such studies vary greatly, the following four descriptions indicate that different models perform better for different applications and calibration can improve model results.

The Thibodeau Study. One of the earliest comprehensive studies was performed by Robert Thibodeau in 1981, which investigated nine software cost models including early versions of PRICE-S and SLIM. The study compared the estimates of the models to actual values for three data bases. From the first data base, Thibodeau gleaned seventeen records from an Air Force data base for information systems software. From the second data base, a military ground systems software data base, he was also able to utilize seventeen values. From the third and final data base, a commercial software data base, Thibodeau found eleven data points. The study showed PRICE-S, when calibrated, averaged within 30% of actual values for military ground systems software. It also showed that SLIM, when calibrated, averaged within 25% of actual values for commercial and information systems software. Thibodeau also noted that when both models were not calibrated, their accuracy's were about five times worse. Although his study did not address recent data or models except PRICE-S and SLIM, Thibodeau did demonstrate the

necessity for model calibration, and that different models were more accurate for different environments (Thibodeau, 1981).

The Illinois Institute of Technology Study. In 1989, the Illinois Institute of Technology performed a similar study using eight Ada language projects and six cost models: SYSTEM-3, which was the predecessor to SYSTEM-4 and SEER, PRICE-S, SASET, SPQR-20, which was the predecessor to CHECKPOINT, and the Ada versions of COCOMO and SoftCost-R. The eight Ada projects were divided into three sub-categories: object-oriented versus structured design, government versus commercial contracts, and command and control versus tools/environment applications (Illinois Institute of Technology (IIT) Research Institute, 1989).

The estimates of the models which were not calibrated to the data base were compared to actual results, and the models were rank-ordered based on how many estimates were within 30% of actual values. SoftCost-Ada and SASET scored highest on overall accuracy; however, models varied in results for sub-categories. For example, SASET and SPQR-20 scored highest for command and control applications. The models were also evaluated for consistency of estimates to within 30% after the mean for the model's estimate was applied. Here, PRICE-S and SYSTEM-3 scored highest; which showed that calibration may enhance the accuracy of these models. The results of this study are consistent with Thibodeau's study in that different models performed better for different (Ada) applications, and calibration can improve model results (Ourada and Ferens, 1991).

Ourada Thesis. In 1991, an AFIT study was conducted which attempted to address four questions:

> 1. Given a credible set of actual DoD data, can the chosen models be calibrated?
> 2. Given a calibrated model, with another set of actual data from the same environment, can the models be validated?
> 3. Given a validated model, if another independent data set from another software environment is used, are the estimates still accurate?
> 4. Is a calibration and validation of a model accurate for only specific areas of application? (Ourada and Ferens, 1991)

Four models, REVIC, SASET, SEER, and COSTMODL were selected for this study based on availability of the models and time constraints. Twenty-eight points were identified that had the same development environment, had data for the actual development effort, had no reused code, and were similar projects when looking at their size. Each model was calibrated on fourteen of the available data points, validated on the remaining fourteen, and analyzed using the same statistical equations that will be identified in Chapter III (Ourada and Ferens, 1991).

The results of this study overwhelmingly showed that these models were not accurate for estimating in the DoD environment. The results; however, should not discourage other researchers from recalibrating these models with a different and perhaps larger data base. It is likely further studies will demonstrate that specific models appear to be accurate for particular environments (Ourada and Ferens, 1991).

The Pentateuch Study. In 1994, SMC requested that the Air Force Institute of Technology (AFIT) calibrate five software cost models: PRICE-S, REVIC, SASET, SEER-SEM, and SLIM to their SWDB. This calibration effort became known as *The*

Pentateuch Study, an effort by five AFIT students to calibrate the above models to the 1994 edition of the SMC SWDB. This data base contained over 2600 entries for DoD-managed programs (Christensen and Ferens, 1995).

Under guidance from MCR and SMC, the database was stratified into the following six categories: Unmanned Space, Military Avionics, Missile, Military Mobile, Military Ground - Signal Processing, and Military Ground - Command and Control (Christensen and Ferens, 1995). SMC decided to divide the data set between calibration and validation as follows:

- If the data points were ≤ 8, then use all the points for calibration only
- If the data points were >8 but ≤ 11, then use 8 for calibration and the rest for validation
- If the data points were ≥ 12, then use 1/3 for validation (Vegas, 1995).

The Pentateuch Study only involved calibration of effort equations or factors; schedule calibration was not done. The primary method for calibrating each of the five models was as follows:

- PRICE-S --- Productivity Factor (PROFAC)
- REVIC --- Coefficient and Exponent; Coefficient Only
- SASET ---Software Type Multiplier; Class Multiplier (alternate)
- SEER-SEM --- Effort Adjustment Factor
- SLIM --- Productivity Index

The sample size for this study ranged from 1 to 11 data points depending on the category and the model. The results showed that only one calibration for REVIC met any

of the statistical criteria that will be explored further in Chapter III. In addition, the equations for the coefficient and exponent of the military ground data set did not make sense from a practical standpoint but rather only represented a "best fit" statistically. None of the calibrations represented a statistically valid estimating equation. PRICE-S showed slightly more promising results in that the calibration produced slightly better results for some data sets but not for others. Not unlike REVIC and PRICE-S, SEER-SEM calibration did not result in a noticeable improvement for most data sets. Calibration for one of the sets, the avionics data set, did result in a substantial improvement but since it was only validated on a single point, it cannot be endorsed as an accurate estimator for other avionics programs (Christensen and Ferens, 1995).

Summary

Computers have become a standard piece of operating equipment in private business and government alike. With the increasing costs to develop software, the field of software cost estimating has rapidly gained the attention of top organizational leaders. This chapter first reviewed research efforts and literature in the area of software cost estimating. Two basic estimating methods, top-down and bottom-up, identified how software costs can be accumulated. Five cost estimating techniques were presented as the ways in which a software cost estimate can be calculated. Second, the basic concepts and background of software cost estimation including the two areas of uncertainty, requirements uncertainty and cost estimating uncertainty, were highlighted to stress the potential for inaccuracies in software cost estimating. Third, a brief look at the evolvement of function points was described to further demonstrate the need for

calibration in this largely untapped area. Lastly, past calibration efforts were examined to provide direction for the research effort and demonstrate the necessity of further calibrations for DoD programs.

The following chapter will describe the calibration procedures for the CHECKPOINT model and the stratification procedures for the SMC SWDB as well as define the statistical measures that will be used to assess the model's estimating accuracy with respect to the baseline or uncalibrated model.

III. Methodology

Overview

The intent of this research effort is to calibrate the CHECKPOINT software cost estimating model to the SMC Software Database (SWDB) and to assess the estimating accuracy of the calibrated model with respect to the baseline or uncalibrated model. This chapter addresses the methods used to calibrate the CHECKPOINT model. First, the CHECKPOINT model and the SMC SWDB are described. Next, the stratification procedures are outlined, assumptions of the data are stated, and a normalization table for the data is given. A step-by-step calibration procedure is then presented for data entry into the model, creating a portfolio, and creating a template. Lastly, statistical measures and methods used in validating the accuracy of the calibrated model are defined and described.

CHECKPOINT, Version 2.3.1

CHECKPOINT is a software cost estimating model that integrates sizing, planning, scheduling, estimating, measurement, risk analysis, value analysis, and technology assessment in a single package (Software Productivity Research, 1993:1-3). This model capitalized and expanded upon the strengths of its predecessor, SPQR-20. For the first time, executives, managers, and their customers, can have a complete view of all tradeoffs between functions, schedules, quality, and costs (Software Productivity Research, 1993:1-3).

CHECKPOINT will be calibrated using projects, profiles, and templates. Each data point will be entered as a project and the data points randomly identified for calibration within each of the eight categories will be saved to a profile. Profiles allow the user to group projects together and serves as the transition between the projects themselves and a template. Eight categories of data or unique applications will exist for calibrating effort for this research, yielding eight profiles. Four of these categories will also be calibrated on schedule, yielding an additional four profiles for a total of twelve. Each profile will then be saved as a template (Zimmerman, 1996). In this effort, the templates are each comprised of only one profile since the data categories are distinct. Templates can comprise numerous profiles when similar categories of projects are assessed for accuracy. The categories will then be compared for trends among them and recommendations will be made.

Data Description

The SMC SWDB was used in calibrating the CHECKPOINT cost model. The SMC SWDB consists of 2,638 records of software development and to a limited extent, maintenance data. The SWDB is an automated (PC based) tool hosted in a windows environment that allows users to easily access and use stored data.

In 1989, SMC contracted MCR, Inc. to develop the SWDB. The purpose of the SWDB was to conduct model calibrations on PRICE-S, REVIC, SASET, SEER-SEM, and SLIM. In addition, SMC wanted to use the database for analogy estimating, estimate verification, and developing cost estimating relationships. The data were collected by mapping other databases and entering information written on forms. Though some of the

data contained in each record are subjective, MCR believes the fields within each record to be consistent across all records. The data were screened and checked against established metrics. The schedule duration field was researched for consistency with size and expanded effort (Stukes, 1996).

The SMC SWDB user's manual gives the following introduction to the database.

> The Space and Missile Center Software Database was developed to access and display data stored in the Space and Missile Systems Center (SMC) Software Database (SWDB). The SWDB was developed under the direction of the USAF Space and Missile Systems Center, with assistance from the Space Systems Cost Analysis Group (SSCAG). (Novak-Ley and Stukes, 1995:3)

The SWDB is a user-friendly program that provides a variety of applications to include a combination of graphical user interface, narrative menus, and help notes. Users can quickly query information along user-defined criteria. The SWDB also allows users to generate a variety of reports on the queried data (Kressin, 1995:34).

Several sources of data comprise the SMC SWDB: government, industry, and other database sources. Government sources include SMC, European Space Agency (ESA), NASA, and Air Force Materiel Command (AFMC) programs. Industry sources include major aerospace companies, suppliers, non-aerospace companies, and model developers. Other data sources include The Aerospace Corporation, SSCAG, General Dynamics, and Jet Propulsion Laboratory (Kressin, 1995:34).

In the process of conducting this effort, a new source of data from a Management Information System (MIS) was identified. A local area contractor provided function point data for effort and schedule from one of their data banks. The SWDB does not currently have many records containing function points so this contribution to their database was

invaluable. This data will be incorporated in the next version of the SWDB and is included as one of the categories in this effort.

Stratification of SWDB

For the purpose of this research and to be consistent with *The Pentateuch Study*, the SWDB was stratified along the following operating environments: Military Mobile, Military - Specific Avionics, Military Ground and Application - Command and Control, Military Ground and Application - Signal Processing, Unmanned Space, Ground in Support of Space, and Missile. These seven subsets of data were restricted to the Computer Software Configuration Item (CSCI) software level.

Two additional categories were also used. The first, Cobol Projects, stemmed from a suggestion from Professor Daniel Ferens. The second, MIS, as described earlier, was collected from a local area DoD contractor and submitted to MCR for inclusion in the next version of the SMC SWDB.

The scope of this research effort will span nine project categories, the first seven to provide continuity with *The Pentateuch Study* and the additional two as previously identified. Each category must have a minimum of eight data points to remain eligible for this study. The rationale for this number stems from last year's efforts in which at least six points were required for each category. After review of the findings, SMC and MCR decided eight points would provide more meaningful results (Stukes, 1996). As will be shown in Chapter IV, one of the project categories, Missile, does not have the required number of data points and therefore was not considered for calibration in this study. The number of available data points will be divided by two; half will be used for calibration and

half for validation. In the case of an odd number of available data points within any project category, the number of points will be divided by two and rounded up so that the number of points calibrated will exceed the number of points validated by one.

Assumptions. In order to conduct a research effort of this magnitude, assumptions had to be made. The first one was that the calibrations would be useful even though some of the data had "holes" in it as will be described in Chapter IV.

The second assumption was that the results obtained from the categories calibrated using the function point method could be compared to the results obtained from the categories using the effort, size, and schedule calibration method. More specifically, categories having function points as their inputs were compared to categories having SLOC as their inputs.

The last assumption was that the data was input correctly into the CHECKPOINT Model so that the calibration results calculated can be considered accurate regarding the data.

Normalization. MCR identified percentages of effort for each phase in software development (Stukes, 1996). However, a local area contractor that provided the MIS data suggested the following normalization percentages for effort and schedule be used in this effort:

Table 2
Normalization of Effort and Schedule

Phase	**Effort**	**Schedule**	**Overall %**
Requirements	.63	1	.06
External Documentation	1.47	2	.14
Internal Documentation	2.1	2	.20
Coding	4.725	5	.45
Integration and Test	1.575	2	.15
Totals:	**10.5**	**12**	**1.00**

Calibration

The data set within each category will be separated into two subsets: A for calibration and B for validation. Each template will be made from the data contained in subset A. For validation, a comparison of the calibrated data against subset B with the uncalibrated data against subset B will be accomplished. Comparison between calibrated data against subset A with uncalibrated data against subset A will not be done since the local area contractor sponsoring this effort felt such a comparison added little to no value to this project.

This research effort will be conducted in three parts: calibration, validation, and a comparison between calibration and validation. As many of the projects contained in the SWDB do not have extensive function point data, the calibrations will be done in one of two ways. Calibrations will first be calculated with function point data when they are

available. If function point data are not available for a given project category, effort, size, and when available, schedule data will be used. This data are given in SLOC and will not be converted to function points prior to completing the calibrations.

During calibration, known model inputs will be adjusted to give accurate outputs. One-half of the database, selected at random will be mathematically adjusted to give outputs as close as possible to the actual outputs reflected in the SWDB (Ourada and Ferens, 1991).

Calibration Procedure

Since CHECKPOINT has not been previously calibrated by or for users outside of a few private organizations, procedures for calibration are not in the model's User Manual. The following list of steps were discovered as a result of a meeting with Mr. John Zimmerman of SPR, Inc. and several meetings with a representative from a local area contractor familiar with calibration procedures of the predecessor to CHECKPOINT, SPQR-20. For further calibrations to be conducted, these steps should be included in the next version of CHECKPOINT's User's Manual.

Step-by-Step Calibration Procedure. To begin calibrating the model, data records must first be input into the model. The data records, known as projects, used in this effort are located in Appendix B. In this appendix, following the application name, each project used in this effort is annotated with its project name in parentheses as it appears on tables 4-15. Only the MIS function points are unique in that the projects used are typed in bold.

Replication Instructions. For the first record only, preliminary background information must be initialized. To begin,

Under SETUP:

1. Go to the field Time Accounting and under 'Project Accounting' identify the accounting hours per business day and the productive hours per project day. In this study, 8.00 for each category was used yielding 251 Business days, 231 Project days, and 1,848 Productive hours.
2. Go to the field, Project Mode and identify type. In this study, 'Measure' was used.
3. Go to the field, Data Entry Level and identify type. In this study, 'Phase' was used.
4. Go to the field, Work Method and identify the time. In this study, 'Months' were used.

Under INPUT, click on 'Required Input' and:

5. Go to the field, Project Description and enter appropriate information paying particular attention to the project's start and end dates.
6. Go to the field, Project Classification and identify the project's nature. In this study, 'New' was assumed.
7. Go to the field, Project Scope and choose from the entries available. 'Programs within a system' was used in this study.
8. Go to the field, Project Class and identify the project. 'External program, developed under military contract' was used in this study.
9. Go to the field, Project Type and identify the project. 'Embedded' was used in this study for all categories except MIS. For this category, 'Interactive Database' was used as the primary and 'Batch' as the secondary.
10. Go to the field, Project Goals and identify accordingly. 'Standard' was used in this study.
11. Go to the field, Project Complexity and identify. 'New problem complexity' was used in this study.

The above 11 steps conclude the entries for preliminary information on data records that will be entered. To enter the first project, continue the above by:

12. Close 'Project Description.'
13. Under Input, click 'Required Input' and then 'Function Sizing.'
14. Enter function points or "0" in each of the five blocks if SLOC are used. Enter the values by pressing 'tab' between each field.
15. Keep pressing 'tab' until the 'Source Code: Source Code Languages' screen appears.
16. Enter appropriate languages by clicking on the 'Choose Languages' box.
17. Enter the appropriate levels for each language and click 'ok.'
18. Enter the SLOC in the 'KLOC' box.

19. Close the 'Source Code' window.
20. Under Input, click 'Measurement Data Entry' and 'Development Effort.'
21. Enter the appropriate 'Schedule Months' and 'Effort Months' values for the five phases: 'Requirements,' 'External Design,' 'Internal Design,' 'Coding,' and 'Integration and Test.' These values are calculated by multiplying the actual data by the normalized values found in Table 3, Normalization of Effort and Schedule.
22. Close this window. Click on 'Requirements' or remember to do so when in the next record since the model does not reset to 'Requirements' automatically.

Once the first project has been entered, subsequent records can be created by recalling the prior one, inputting the new values, and saving it with a new name. To open an existing project:

1. Under 'File,' click 'Open' and then 'Project.'
2. Highlight desired project and click 'ok.'
3. Click 'ok' on the 'Open Version.'
4. Continue with steps 12-22.

Once all projects have been entered, a portfolio can be created to group several projects together:

1. Under 'File,' click 'New' and 'Portfolio.'
2. Highlight the records desired in the Portfolio. Click 'ok.'
3. Under 'File,' 'Save as' and close.

The last step in the calibration process is to create a template. This can be done by:

1. Under 'File,' click 'New' and 'Template.'
2. In the 'New Template' window, leave all default values as they appear on the screen and click 'ok.'

Validation

"Calibration without validation is meaningless" (Christensen, 1996). During validation, the other half of the SWDB will be used as input data but the calibrated model parameters will not be changed (Ourada and Ferens, 1991). A comparison between the

actual values obtained for effort and schedule for each project in the validation data set will be made with the estimates obtained from the calibrated template. Five measures will be used to validate the estimating accuracy of the calibrated and uncalibrated models: the Magnitude of Relative Error (MRE), the Mean Magnitude of Relative Error (MMRE), the Root Mean Square Error (RMS), the Relative Root Mean Square Error (RRMS), and the Prediction at Level *k/n* (PRED(*l*)) (Conte, Dunsmore, and Shen, 1986:172). In addition, the Wilcoxon Signed-Rank Test will be used to test for bias associated with the calibrated model. They are now further explained in the following paragraphs:

Magnitude of Relative Error (MRE). The MRE reflects the degree of estimating error in a particular estimate. This measure was calculated for each software project contained in the validation data sets. MRE is defined by the equation:

$$MRE = \left| \frac{E_{act} - E_{est}}{E_{act}} \right| \tag{1}$$

where

E_{act} = actual normalized total development effort reported in the SMC SWDB

E_{est} = estimated total development effort as reported by CHECKPOINT

Note that as the MRE becomes smaller, the estimate is said to become more accurate and reflect less error (Conte, Dunsmore, and Shen, 1986:172).

Mean Magnitude of Relative Error (MMRE). The MMRE reflects the average degree of estimating error produced by a set of estimates. An MMRE was calculated for

each validation data set of both calibrated and uncalibrated models. The MMRE is defined by the equation:

$$MMRE = \frac{1}{n} * \sum_{i=1}^{n} MRE_i \tag{2}$$

where

n = total number of records (software projects) in a particular data set

i = the ith record in a particular data set

MRE = Magnitude of Relative Error

Note that the smaller the MMRE the better the model produces on average a set of estimates. For the model to be acceptable, MMRE should be less than or equal to 0.25. (Conte, Dunsmore, and Shen, 1986:172).

Root Mean Square Error (RMS). The smaller the value of RMS, the better the model's ability to forecast actual performance. The RMS is defined by the equation:

$$RMS = \sqrt{\frac{1}{n}\sum_{n=1}^{n}(E_{act} - E_{est})^2} \tag{3}$$

where

n = total number of records (software projects) in a particular data set

E_{act} = actual normalized total development effort reported in the SMC SWDB

E_{est} = estimated total development effort as reported by CHECKPOINT

Relative Root Mean Square Error (RRMS). The smaller the value of RRMS, the better the model's ability to forecast actual performance. The RRMS is defined by the equation:

$$RRMS = \frac{RMS}{\frac{1}{n}\sum_{n=1}^{n} E_{act}} \tag{4}$$

where

n = total number of records (software projects) in a particular data set

E_{act} = actual normalized total development effort reported in the SMC SWDB

RMS = Root Mean Square Error

For RRMS, an acceptable model will give a value of RRMS that is less than 0.25 (Conte, Dunsmore, and Shen, 1986:172).

Prediction at Level *k/n* (PRED(*l*)). This measure is sometimes known as the percentage method and is used frequently in validating and reporting the predictive accuracy of a model. It basically reflects the percentage of project estimates in a given data set that fall within a predefined percentage of their actual values. This measure is defined as:

$$PRED(l) = \frac{k}{n} \tag{5}$$

where

k = number of software projects in a particular data set of n projects whose

$$MRE \leq \frac{k}{n}$$

n = total number of software projects in a particular data set

As an example, "if PRED(0.25) = 0.83, then 83% of the predicted values fall within 25% of their actual values. To establish the model accuracy, 75% of the predictions must fall within 25% of the actual values, or PRED (0.25) >= 0.75" (Conte, Dunsmore, and Shen, 1986:173).

Wilcoxon Signed-Rank Test. The Wilcoxon signed-rank test is a non-parametric test and was used to test for bias in the distributions of the estimate and actual observations. The absolute value of the differences were ranked from the least to the greatest. If the data were truly unbiased, one would expect that just as many positive differences would occur as would negative differences, therefore, the number of positive and negative differences would sum to zero. As such, the ranking was then partitioned into rankings of positive (T+) and negative (T-) differences. "Sizable differences in the sums of the ranks assigned to the positive and negative differences would provide evidence to indicate a shift in location between the distributions" (Mendenhall, Wackerly, and Scheaffer, 1990:680). If there exists a statistically significant difference in the sums of the ranks assigned to the positive and negative differences, one would conclude that the estimate observations, when compared to the actuals, are biased toward being either high or low.

Summary

This chapter focused on the methods used to calibrate the CHECKPOINT model. The stratification procedures followed to glean data from the SMC SWDB were presented along with assumptions of the data and the normalization table. Step-by-step calibration procedures spanning data entry, portfolio creation, and template creation were outlined. Statistical measures and methods used in validating the accuracy of the calibrated model were defined and described. The results of the calibration on the eight categories of data, are found and analyzed in Chapter IV.

IV. Findings and Analysis

Overview

This chapter presents the analysis of the data and gives the results. The chapter describes the assumptions made about missing data, adjustments made to the data, results of the calibrations for each of the eight categories, and a comparison of categories calibrated on function points with those calibrated on effort and SLOC as well as those calibrated on effort, SLOC, and schedule. A contrast between effort and schedule results will also be presented. In addition, the dramatic increase in accuracy evident in the comparison of uncalibrated model estimates against calibrated model estimates will be highlighted.

The Data

The SWDB contains 2,638 records. Each record represents one data point. This data was stratified into the following eight categories for this calibration:

- Military Mobile
- Military - Specific Avionics
- Military Ground and Application -- Command & Control
- Military Ground and Application -- Signal Processing
- Unmanned Space
- Ground in Support of Space
- Cobol Projects
- Missile

A search was accomplished by CSCI and the following data were requested for each category: Schedule Months, Total Effort, Normal Effort Size (otherwise referred to as SLOC) and Function Point Information which included: External Inputs, External Outputs, External Inquires, Internal Files, and External Interfaces.

The ninth category, MIS, with seventy data points, fifteen of which have the required information for this study, was collected from a local area DoD contractor and submitted to MCR for inclusion in the next version of the SMC SWDB. This project category was also used in this study.

Identification. Three categories with information on function points, MIS, Military Mobile, and Military - Specific Avionics, were calibrated using the function point data. The remaining five categories, Military Ground and Application - Command and Control, Military Ground and Application - Signal Processing, Unmanned Space, Ground in Support of Space, and Cobol Projects, were calibrated using effort and size (SLOC). Four categories, MIS, Unmanned Space, Ground in Support of Space, and Cobol Projects were also calibrated on schedule data. The Missile category was not calibrated since only five points were identified from the search. In Chapter I, the guidelines for stratification were presented. If a category had less than eight points, it would not be calibrated.

Further stratification of the data on language enabled four categories, MIS, Military Mobile, Ground in Support of Space, and Cobol Projects to be calibrated on a specific language. MIS and Cobol Projects were calibrated on COBOL and Military Mobile and Ground in Support of Space were calibrated on Ada.

The following table provides a summary of each category, the calibration method used, the number of points available in the category, the language, the number of points available per language, the language calibrated on (if any), and a status column denoting which method; either function points or effort, SLOC, and schedule was used for calibration and how many data points were used.

The numbers in parentheses denote the amount of data points that could have been gleaned if the requirements for each point were not binding. For example, in the Military Mobile category, eight points were identified, five of which were Ada. Three more Ada points were found but their function point information was not complete.

Table 3
Summary of Calibration Categories, Methods, Languages, and Number of Data Points

Category	Calibration Method	# of Points	Language	# of Points	Language (if any)	Status
MIS	**Function Points (including Sch)**	**13**	**COBOL**	**13**	**COBOL**	**FP (13); Sch**
Military Mobile	**Function Points**	**8**	**Ada**	**5 (3)**	**Ada**	**FP (8)**
Military - Specific Avionics	**Function Points**	**8**	**Ada**	**3**	**None**	**FP (8)**
			Jovial	**2**		
			FORTRAN	**1**		
			Mix	**1**		
			Other	**1**		
Mil Ground - C&C	**Effort/SLOC**	**13**	**FORTRAN**	**2**	**None**	**Effort/SLOC (13)**
			C	**1**		
			Mix	**2**		
			Unknown	**8**		
Mil Ground - S&P	**Effort/SLOC**	**20**	**Ada**	**1**	**None**	**Effort/SLOC (20)**
			Pascal	**1**		
			Unknown	**18**		
Unmanned Space	**Effort/SLOC/Sch**	**17**	**Ada**	**6**	**None**	**Effort/SLOC/ Sch (11)**
			C	**3 (3)**		
			Assembly	**2 (3)**		
Ground in Support of Space	**Effort/SLOC/Sch**	**47**	**Ada**	**7 (12)**	**Ada**	**Effort/SLOC/ Sch (8)**
			Assembly	**4 (3)**		
			FORTRAN	**5 (4)**		
			Pascal	**4 (1)**		
			Other	**6**		
			C	**1**		
Cobol Projects	**Effort/SLOC/Sch**	**13**	**COBOL**	**7**	**COBOL**	**Effort/SLOC/ Sch (8)**
			Mix	**6**		
Missile	**Effort/SLOC**	**5**	**Assembly**	**4**	**None**	**< 8 data pts**
			Jovial	**1**		

In the Unmanned Space category, three data points were identified for the C language and two for Assembly. However, three more points for each language could have been used since these points were mostly comprised of either C or Assembly in conjunction with a small percentage of other languages. It was determined that this category would best be calibrated by not stratifying on language since none of the three languages, Ada with six points, C also with six points, nor Assembly with five points had enough data as previously defined by the stratification procedures. The solid eleven data points were used for a more robust sample size rather than the eleven plus the non-complete six.

In the Ground in Support of Space category, a total of eight points were stratified on the Ada language. A possible 47 data points were identified; 27 of which were complete. Of the 27 complete points, the only language that had almost sufficient data was Ada with seven. Calibration of this category could be done by adding one of the twelve identified points included earlier when considering points that were not completely written in the Ada language. It was decided to include only one such point rather than all twelve to preserve the sample as much as possible. Similarly, it was determined that calibrating this category on the FORTRAN language with five authentic points and four "almost" authentic points would not be worthwhile.

The Results

The following twelve tables provide the calibration and validation information for the eight categories. All eight categories were calibrated on effort and are presented in

tables 4-11. Four categories, MIS, Unmanned Space, Ground in Support of Space, and Cobol Projects were also calibrated on schedule and are presented in tables 12-15.

Each category begins with a description of its calibration method, the language it was stratified on (if any), and the number of points used for calibration and validation. The table for each category lists the project names as they were displayed in the SWDB in column 1. Columns 2-5 list the actual effort, the estimated effort, the MRE, and the Wilcoxon for validation. Columns 6-8 list the effort, MRE, and Wilcoxon for calibration. Below these columns are the MMRE, RMS, RRMS, and PRED*(l)* values for validation, (on the left side) and calibration, (on the right side). Following the calculations is a graph depicting the relationship among calibrated effort, estimated effort, and actual effort for each project.

Below the graph, further analysis for the MMRE, the RRMS, the PRED*(l)*, and the Wilcoxon Signed-Rank Test is provided, first for the calibrated model and then for the uncalibrated model. A comparison between the calibrated and uncalibrated results is then presented. Since the MMRE and the RRMS incorporate the MRE and RMS values, respectively, further analysis addressing each value in the categories of the latter two statistics is not provided. The smaller the value of MRE, the better, indicating that the model is predicting accurately. In addition, the smaller the value of the RMS, the better the model is at estimating.

MIS (Effort). This category was calibrated using function points to include schedule data. It was further stratified on the COBOL language and thirteen points were used, seven for calibration and six for validation.

The following table provides summary statistical information:

Table 4
MIS (Effort)

Project	Effort (Actual)	Effort (Uncalibrated)	MRE	Wilcoxon	Effort (Calibrated)	MRE	Wilcoxon
MIS 4	**10**	**22.63**	**1.263**	**-12.63**	**10.49**	**0.049**	**-0.49**
MIS 5	**6**	**15.89**	**1.648**	**-9.89**	**6.37**	**0.062**	**-0.37**
MIS 6	**30**	**36.68**	**0.223**	**-6.68**	**30.27**	**0.009**	**-0.27**
MIS 7	**132**	**135.10**	**0.023**	**-3.10**	**132.14**	**0.001**	**-0.14**
MIS 8	**296**	**305.07**	**0.031**	**-9.07**	**296.43**	**0.001**	**-0.43**
MIS 9	**673**	**716.00**	**0.064**	**-43.00**	**674.60**	**0.002**	**-1.60**
		MMRE	**0.542**		**MMRE**	**0.018**	
		RMS	**19.334**		**RMS**	**0.678**	
		RRMS	**0.101**		**RRMS**	**0.004**	
		PRED(l)	**0.667**		**PRED(l)**	**1.000**	

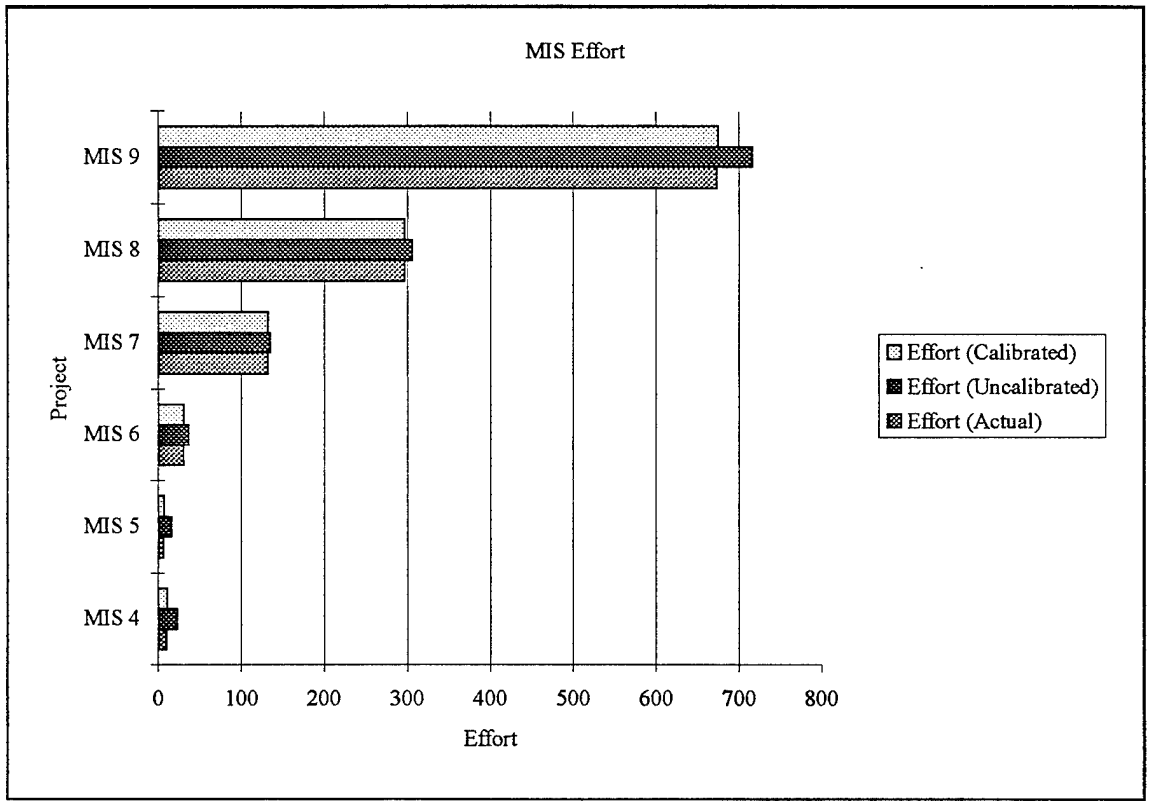

Figure 1. MIS (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort

Calibrated Model. As the MMRE for the calibrated model was approximately .018, it was highly acceptable according to Conte's criteria which define an acceptable model as one with an MMRE less than or equal to 0.25 (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also echoed this high level of acceptability. Its value was approximately .004, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 100% of the predicted values fell within 25% of their actual values. This value greatly exceeded the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a slight negative bias.

For this category, the three statistical tests, MMRE, RRMS, and PRED(*l*) demonstrated the model was highly acceptable.

Uncalibrated Model. As the MMRE for the uncalibrated model was approximately .542, it was unacceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, however, did not support this hypothesis since its value was approximately .101, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, approximately 67% of the predicted values fell within 25% of their actual values. This value did not exceed the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a negative bias.

For this category, the RRMS showed the model was acceptable. However, two statistical tests, MMRE and PRED(*l*), did not support this hypothesis.

Comparison Between the Calibrated Model and the Uncalibrated Model. The MMRE for the calibrated model decreased by approximately .524, making it 96.71% more accurate. The RRMS decreased by approximately .097, making it 95.91% more accurate. The PRED(*l*) increased by approximately .34, making it a more than acceptable estimator for model accuracy.

Military Mobile (Effort). This category was calibrated using function points. It was further stratified on the Ada language and eight points were used, four for calibration and four for validation.

The following table provides summary statistical information:

Table 5
Military Mobile (Effort)

Project	Effort (Actual)	Effort (Uncalibrated)	MRE	Wilcoxon	Effort (Calibrated)	MRE	Wilcoxon
MM5	**920**	**1080.61**	**0.175**	**-160.61**	**941.88**	**0.024**	**-21.88**
MM6	**211.1**	**362.06**	**0.715**	**-150.96**	**231.91**	**0.099**	**-20.81**
MM7	**179**	**329.71**	**0.842**	**-150.71**	**199.73**	**0.116**	**-20.73**
MM8	**15**	**72.04**	**3.803**	**-57.04**	**22.93**	**0.529**	**-7.93**
		MMRE	**1.384**		**MMRE**	**0.192**	
		RMS	**136.521**		**RMS**	**18.738**	
		RRMS	**0.412**		**RRMS**	**0.057**	
		PRED(l)	**0.250**		**PRED(l)**	**0.750**	

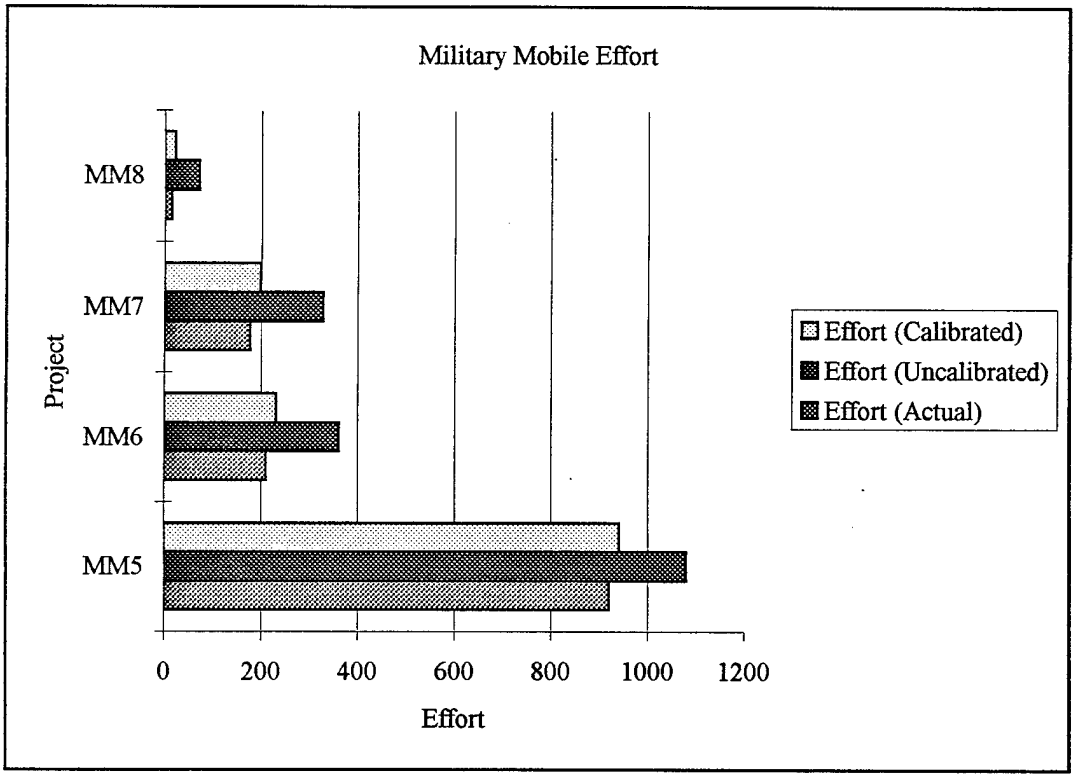

Figure 2. Military Mobile (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort

Calibrated Model. As the MMRE for the calibrated model was approximately .192, it was more than acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .057, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 75% of the predicted values fell within 25% of their actual values. This value equaled the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a negative bias in the data.

As in the preceding category, MIS (effort), the three statistical tests, MMRE, RRMS, and PRED(*l*), demonstrated the model was highly acceptable for Military Mobile (effort).

Uncalibrated Model. As the MMRE for the uncalibrated model was approximately 1.38, it was unacceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .412, a value much larger than the acceptable 0.25.

When analyzing the PRED (0.25) value, 25% of the predicted values fell within 25% of their actual values. This value was far less than the desired percentage of 75% to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a negative bias in the data.

The three statistical tests, MMRE, RRMS, and PRED(*l*), demonstrated the uncalibrated model was unacceptable for Military Mobile (effort).

Comparison Between the Calibrated Model and the Uncalibrated Model. The MMRE decreased by approximately 1.19, making it 86.14% more accurate. The RRMS decreased approximately .356, making it 86.27% more accurate. The PRED(*l*) increased by .5, making it an acceptable estimator for model accuracy.

Military - Specific Avionics (Effort). This category was calibrated using function points. It was not further stratified on language. Eight points were used, four for calibration and four for validation.

The following table provides summary statistical information:

Table 6
Military - Specific Avionics (Effort)

Project	Effort (Actual)	Effort (Uncalibrated)	MRE	Wilcoxon	Effort (Calibrated)	MRE	Wilcoxon
Milspav 5	**259**	**510.39**	**0.971**	**-251.39**	**296.02**	**0.143**	**-37.02**
Milspav 6	**37**	**119.67**	**2.234**	**-82.67**	**51.84**	**0.401**	**-14.84**
Milspav 7	**68**	**64.86**	**0.046**	**3.14**	**64.66**	**0.049**	**3.34**
Milspav 8	**409**	**415.21**	**0.015**	**-6.21**	**393.86**	**0.037**	**15.14**
		MMRE	**0.817**		**MMRE**	**0.158**	
		RMS	**132.363**		**RMS**	**21.396**	
		RRMS	**0.685**		**RRMS**	**0.111**	
		PRED(l)	**0.500**		**PRED(l)**	**0.750**	

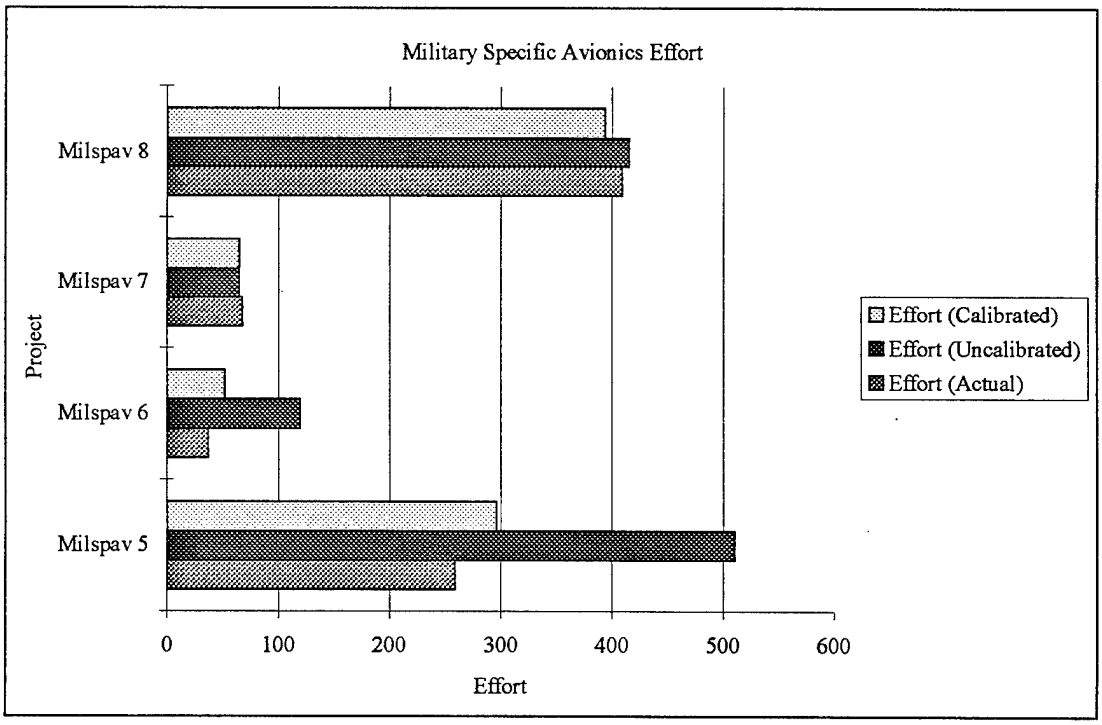

Figure 3. Military - Specific Avionics (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort

Calibrated Model. As the MMRE for the calibrated model was approximately .157, it was more than acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .111, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 75% of the predicted values fell within 25% of their actual values. This value equaled the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a negative bias in the data.

As in the preceding categories, MIS (effort) and Military Mobile (effort), the three statistical tests, MMRE, RRMS, and PRED(*l*), demonstrated the model was highly acceptable for Military - Specific Avionics (effort).

Uncalibrated Model. As the MMRE for the uncalibrated model was approximately .817, it was unacceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .685, a value much larger than the acceptable 0.25.

When analyzing the PRED (0.25) value, 50% of the predicted values fell within 25% of their actual values. This value was less than the desired percentage of 75% to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a negative bias in the data.

As in the preceding category, Military Mobile (effort), the three statistical tests, MMRE, RRMS, and PRED(*l*), demonstrated the model was unacceptable for Military - Specific Avionics (effort).

Comparison Between the Calibrated Model and the Uncalibrated Model. The MMRE decreased by approximately .659, making it 80.71% more accurate. The RRMS decreased approximately .574, making it 83.84% more accurate. The PRED(*l*) increased by .25, making it an acceptable estimator for model accuracy.

Military Ground and Application - Command & Control (Effort). This category was calibrated using effort and SLOC data. It was not further stratified on language. Thirteen points were used, seven for calibration and six for validation.

The following table provides summary statistical information:

Table 7
Military Ground and Application- Command & Control (Effort)

Project	Effort (Actual)	Effort (Uncalibrated)	MRE	Wilcoxon	Effort (Calibrated)	MRE	Wilcoxon
MGCC 4	**656**	**656.00**	**0.000**	**0.00**	**656.01**	**0.000**	**0.00**
MGCC 5	**95**	**173.54**	**0.827**	**-78.54**	**173.54**	**0.827**	**-78.54**
MGCC 6	**139**	**139.00**	**0.000**	**0.00**	**139.00**	**0.000**	**0.00**
MGCC 7	**322**	**322.02**	**0.000**	**-0.02**	**322.02**	**0.000**	**-0.02**
MGCC 8	**101**	**129.10**	**0.278**	**-28.10**	**129.10**	**0.278**	**-28.10**
MGCC 9	**100**	**95.00**	**0.050**	**5.00**	**95.00**	**0.050**	**5.00**
		MMRE	**0.193**		**MMRE**	**0.165**	
		RMS	**34.115**		**RMS**	**31.585**	
		RRMS	**0.145**		**RRMS**	**0.156**	
		PRED(l)	**0.500**		**PRED(l)**	**0.500**	

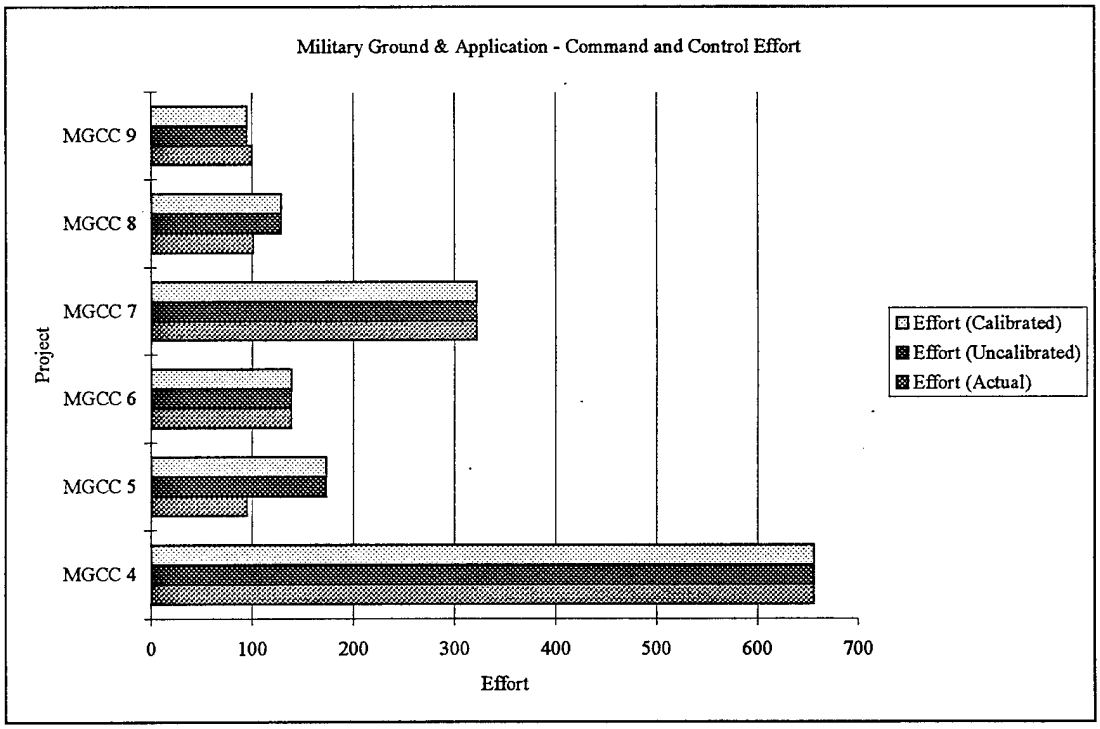

Figure 4. Military Ground and Application - Command and Control (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort

Calibrated Model. As the MMRE for the calibrated model was approximately .165, it was more than acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .156, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 50% of the predicted values fell within 25% of their actual values. This value was below the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a negative bias.

Two out of the three statistical tests, MMRE and RRMS demonstrated the model was highly acceptable for this category. However, the PRED(*l*) value did not support this hypothesis.

Uncalibrated Model. As the MMRE for the uncalibrated model was approximately .193, it was more than acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .145, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 50% of the predicted values fell within 25% of their actual values. This value was below the desired percentage of 75% to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a negative bias.

Two out of the three statistical tests, MMRE and RRMS demonstrated the model was highly acceptable for this category. However, the PRED(*l*) value did not support this hypothesis.

Comparison Between the Calibrated Model and the Uncalibrated Model. The MMRE decreased by approximately .028, making it 14.29% more accurate. The RRMS increased by approximately :012, still falling well below the .25 criteria. The PRED(*l*) did not change.

Military Ground and Application - Signal Processing (Effort). This category was calibrated using effort and SLOC data. It was not further stratified on language. Twenty points were used, ten for calibration and ten for validation.

The following table provides summary statistical information:

Table 8
Military Ground and Application -- Signal Processing (Effort)

Project	Effort (Actual)	Effort (Uncalibrated)	MRE	Wilcoxon	Effort (Calibrated)	MRE	Wilcoxon
Milgrs 20	**113**	**84.80**	**0.250**	**28.20**	**84.80**	**0.250**	**28.20**
Milgrs 1	**134**	**100.54**	**0.250**	**33.46**	**100.54**	**0.250**	**33.46**
Milgrs 2	**165**	**156.80**	**0.050**	**8.20**	**156.80**	**0.050**	**8.20**
Milgrs 3	**13**	**12.35**	**0.050**	**0.65**	**12.35**	**0.050**	**0.65**
Milgrs 4	**738**	**701.10**	**0.050**	**36.90**	**701.10**	**0.050**	**36.90**
Milgrs 5	**192**	**182.40**	**0.050**	**9.60**	**182.40**	**0.050**	**9.60**
Milgrs 6	**278**	**264.10**	**0.050**	**13.90**	**264.10**	**0.050**	**13.90**
Milgrs 7	**645**	**612.75**	**0.050**	**32.25**	**612.75**	**0.050**	**32.25**
Milgrs 8	**228**	**216.60**	**0.050**	**11.40**	**216.60**	**0.050**	**11.40**
Milgrs 9	**264**	**250.76**	**0.050**	**13.24**	**250.76**	**0.050**	**13.24**
		MMRE	**0.090**		**MMRE**	**0.090**	
		RMS	**22.304**		**RMS**	**22.304**	
		RRMS	**0.081**		**RRMS**	**0.081**	
		PRED(l)	**1.000**		**PRED(l)**	**1.000**	

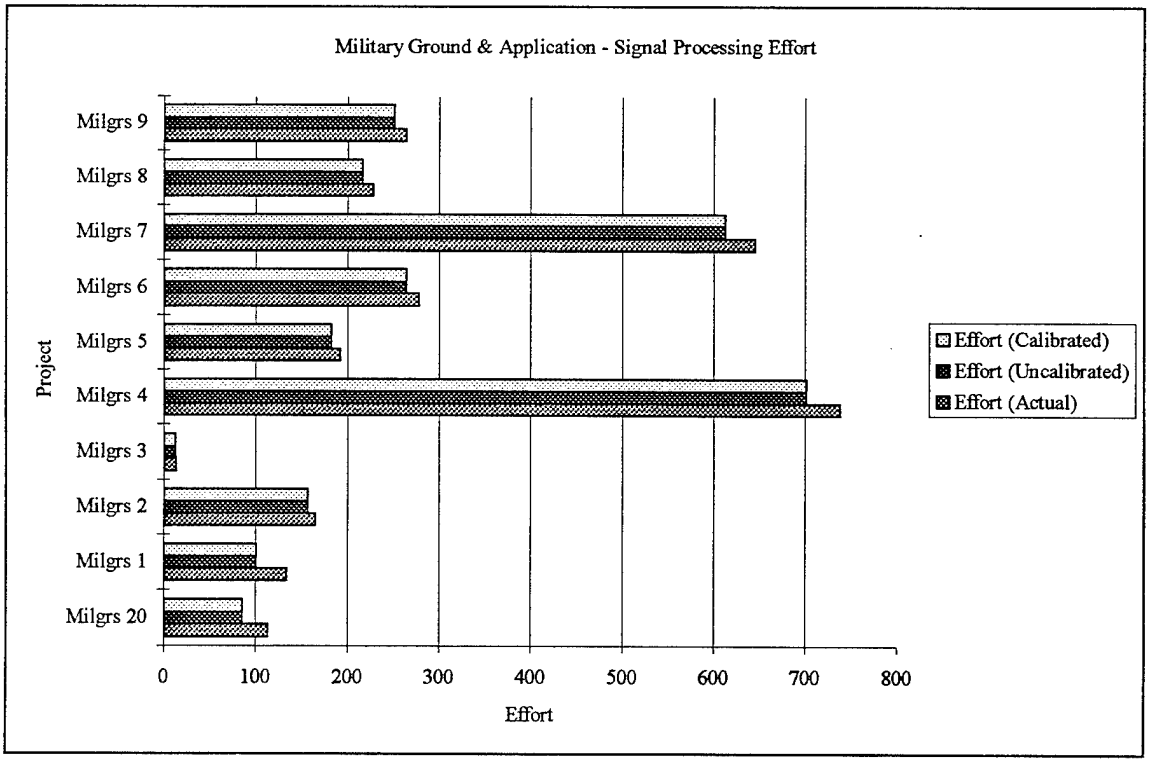

Figure 5. Military Ground and Application - Signal Processing (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort

Calibrated Model. As the MMRE for the calibrated model was approximately .090, it was highly acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .081, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 100% of the predicted values fell within 25% of their actual values. This value far exceeded the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a positive bias.

The three statistical tests, MMRE, RRMS, and PRED*(l)*, demonstrated the model was highly acceptable for this category.

Uncalibrated Model. As the MMRE for the uncalibrated model was approximately .090, it was highly acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .081, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 100% of the predicted values fell within 25% of their actual values. This value far exceeded the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a positive bias.

The three statistical tests, MMRE, RRMS, and PRED*(l)*, demonstrated the model was highly acceptable for this category.

Comparison Between the Calibrated Model and the Uncalibrated Model. There was no change in the MMRE, RRMS, or the PRED*(l)* statistics. These statistics were all acceptable before being calibrated. Since the estimates were the same for both the uncalibrated and calibrated model, calibration did not increase the estimating accuracy.

Unmanned Space (Effort). This category was calibrated using effort, SLOC, and schedule data. It was not further stratified on language. Eleven points were used, six for calibration and five for validation.

The following table provides summary statistical information:

Table 9
Unmanned Space (Effort)

Project	Effort (Actual)	Effort (Uncalibrated)	MRE	Wilcoxon	Effort (Calibrated)	MRE	Wilcoxon
SigProc 7	**200.00**	**191.90**	**0.041**	**8.10**	**191.90**	**0.041**	**8.10**
SigProc 8	**410.00**	**389.50**	**0.050**	**20.50**	**389.50**	**0.050**	**20.50**
SigProc 9	**321.70**	**305.64**	**0.050**	**16.06**	**305.64**	**0.050**	**16.06**
SigProc 10	**321.70**	**305.64**	**0.050**	**16.06**	**305.64**	**0.050**	**16.06**
SigProc 11	**321.70**	**305.64**	**0.050**	**16.06**	**305.64**	**0.050**	**16.06**
		MMRE	**0.048**		**MMRE**	**0.040**	
		RMS	**15.872**		**RMS**	**14.489**	
		RRMS	**0.050**		**RRMS**	**0.055**	
		PRED(l)	**1.000**		**PRED(l)**	**1.000**	

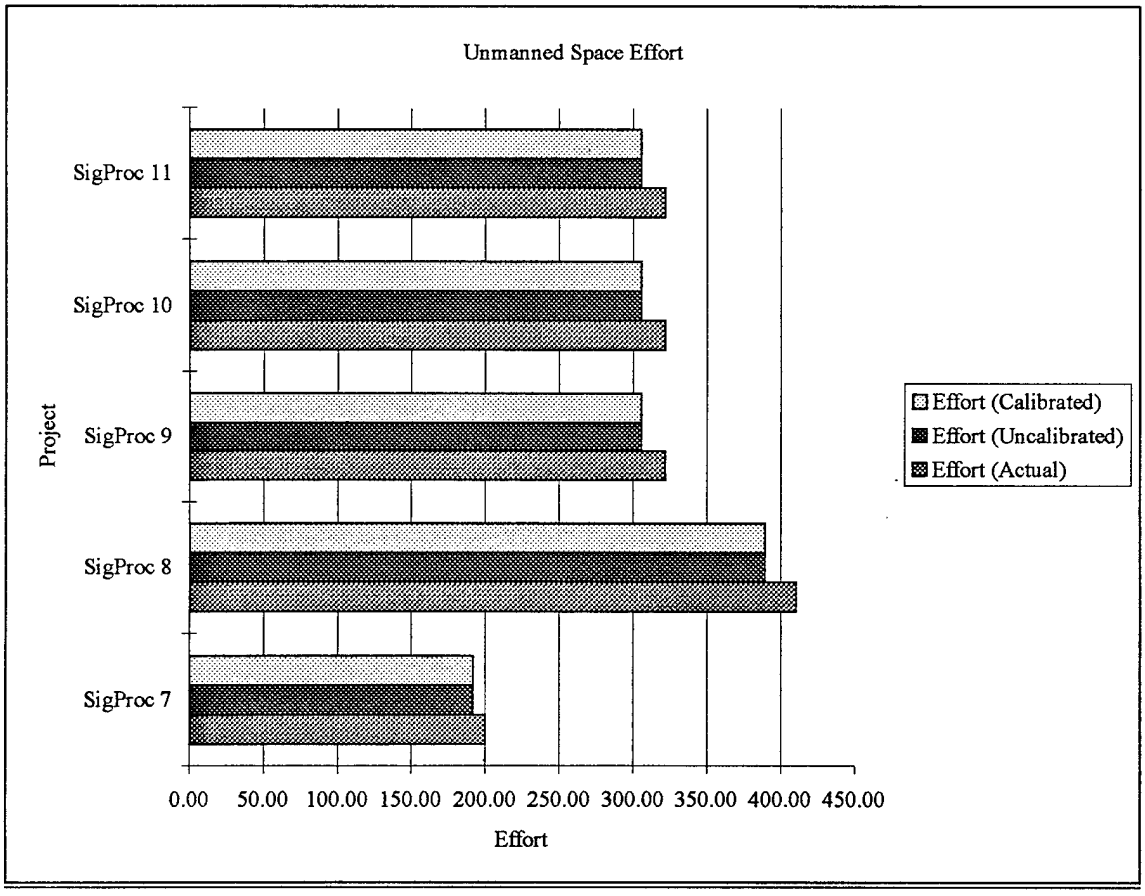

Figure 6. Unmanned Space (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort

Calibrated Model. As the MMRE for the calibrated model was approximately .040, it was highly acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .055, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 100% of the predicted values fell within 25% of their actual values. This value far exceeded the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a positive bias.

As in the previous categories, the three statistical tests, MMRE, RRMS, and PRED(*l*), demonstrated the model was acceptable for this category.

Uncalibrated Model. As the MMRE for the uncalibrated model was approximately .048, it was highly acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .050, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 100% of the predicted values fell within 25% of their actual values. This value far exceeded the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a positive bias.

The three statistical tests, MMRE, RRMS, and PRED(*l*), demonstrated the model was acceptable for this category.

Comparison Between the Calibrated Model and the Uncalibrated Model. The MMRE decreased by approximately .008, making it 16.67% more accurate. The RRMS increased by approximately .005, still falling well below the .25 criteria. The PRED*(l)* did not change but it was already at 100% for the uncalibrated model.

Ground in Support of Space (Effort). This category was calibrated using effort, SLOC, and schedule data. It was further stratified on the Ada language. Eight points were used, four for calibration and four for validation.

The following table provides summary statistical information:

Table 10
Ground in Support of Space (Effort)

Project	Effort (Actual)	Effort (Uncalibrated)	MRE	Wilcoxon	Effort (Calibrated)	MRE	Wilcoxon
5	**15**	**14.30**	**0.047**	**0.70**	**14.30**	**0.047**	**0.70**
6	**20.1**	**19.00**	**0.055**	**1.10**	**19.00**	**0.055**	**1.10**
7	**71**	**67.50**	**0.049**	**3.50**	**67.50**	**0.049**	**3.50**
8	**74.3**	**70.60**	**0.050**	**3.70**	**70.60**	**0.050**	**3.70**
		MMRE	**0.050**		**MMRE**	**0.050**	
		RMS	**2.629**		**RMS**	**2.629**	
		RRMS	**0.058**		**RRMS**	**0.058**	
		PRED(l)	**1.000**		**PRED(l)**	**1.000**	

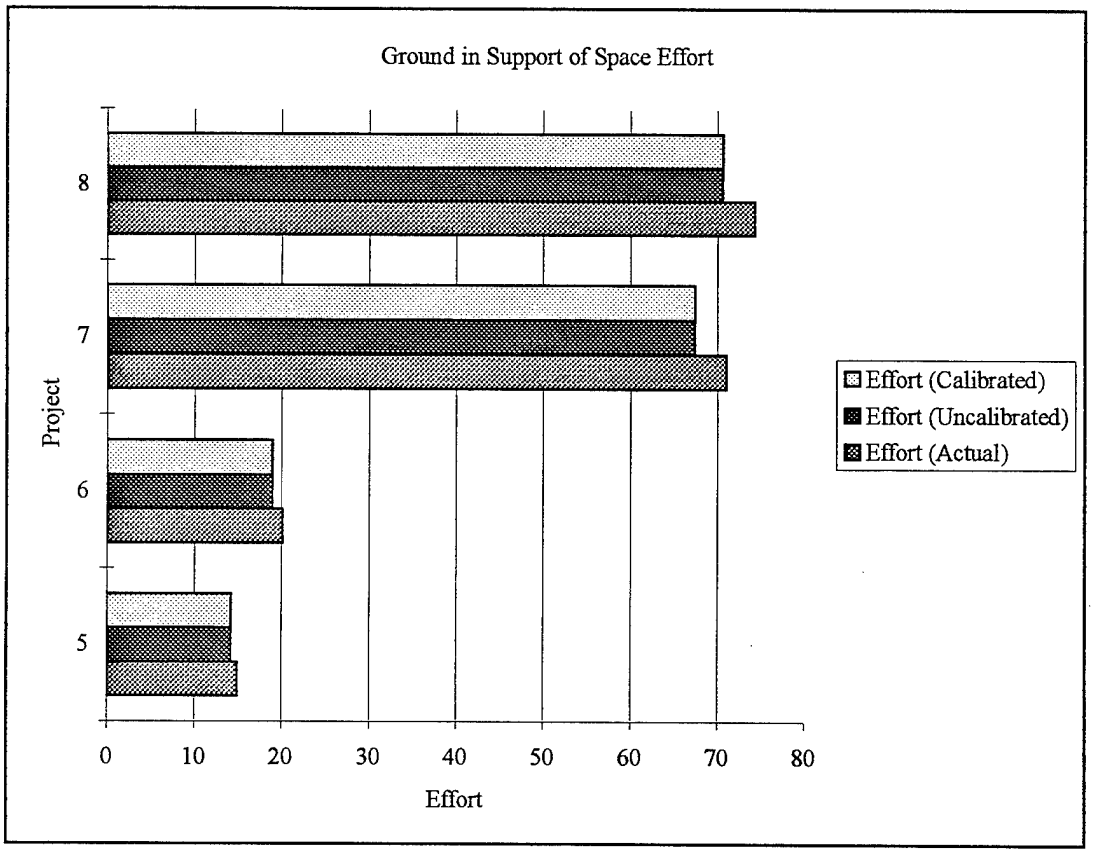

Figure 7. Ground in Support of Space (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort

Calibrated Model. As the MMRE for the calibrated model was approximately .050, the calibrated model was highly acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .058, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 100% of the predicted values fell within 25% of their actual values. This value far exceeded the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a slight positive bias.

The three statistical tests, MMRE, RRMS, and PRED*(l)*, demonstrated the model was highly acceptable for this category.

Uncalibrated Model. As the MMRE for the uncalibrated model was approximately .050, it was highly acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .058, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 100% of the predicted values fell within 25% of their actual values. This value far exceeded the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a slight positive bias.

The three statistical tests, MMRE, RRMS, and PRED*(l)*, demonstrated the model was highly acceptable for this category.

Comparison Between the Calibrated Model and the Uncalibrated Model. There was no change in the MMRE, RRMS, or the PRED*(l)* statistics. These statistics were all acceptable before being calibrated. Since the estimates were the same for both the uncalibrated and calibrated model, calibration did not improve estimating accuracy.

Cobol Projects (Effort). This category was calibrated using effort, SLOC, and schedule data. It was further stratified on the COBOL language. Eight points were used, four for calibration and four for validation.

The following table provides summary statistical information:

Table 11
Cobol Projects (Effort)

Project	Effort (Actual)	Effort (Uncalibrated)	MRE	Wilcoxon	Effort (Calibrated)	MRE	Wilcoxon
5	**652**	**619.42**	**0.050**	**32.58**	**619.42**	**0.050**	**32.58**
6	**438**	**417.10**	**0.048**	**20.90**	**417.10**	**0.048**	**20.90**
7	**358**	**340.10**	**0.050**	**17.90**	**340.10**	**0.050**	**17.90**
8	**299**	**284.70**	**0.048**	**14.30**	**284.70**	**0.048**	**14.30**
		MMRE	**0.049**		**MMRE**	**0.049**	
		RMS	**22.490**		**RMS**	**22.490**	
		RRMS	**0.051**		**RRMS**	**0.051**	
		PRED(l)	**1.000**		**PRED(l)**	**1.000**	

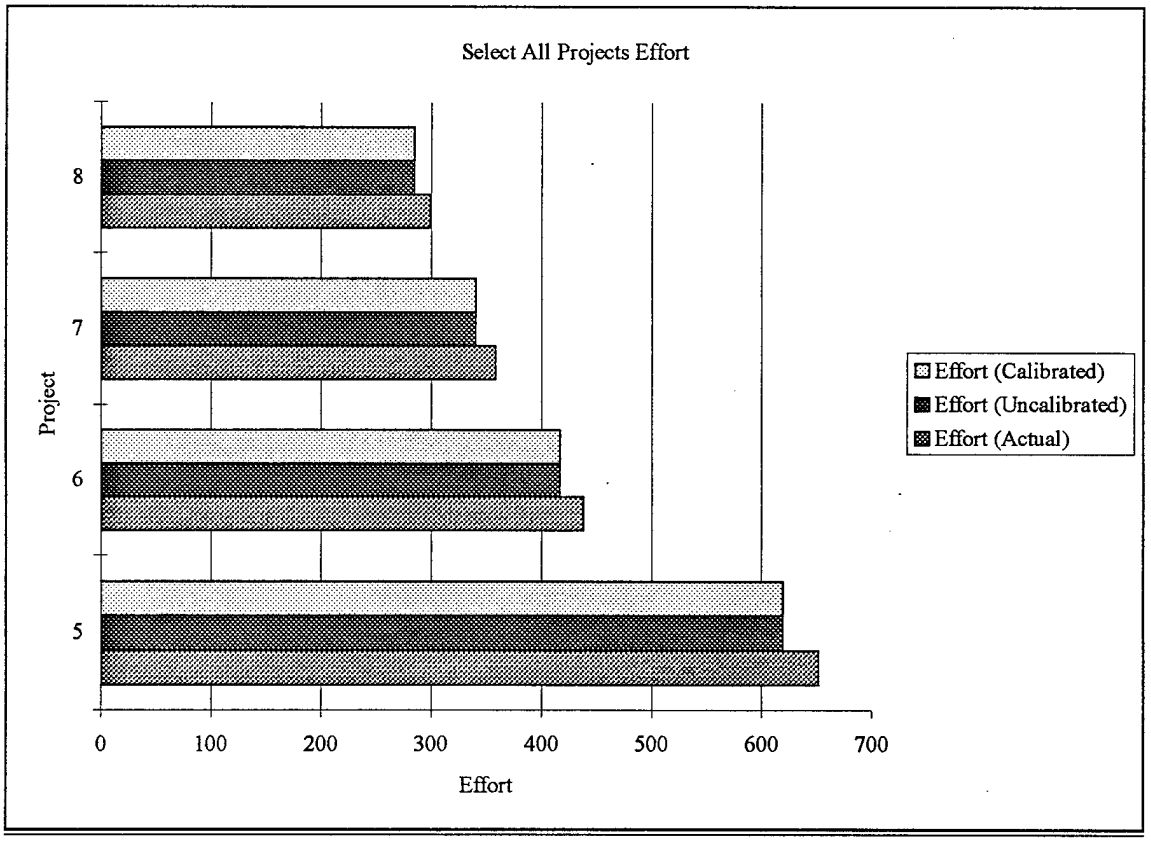

Figure 8. Cobol Projects (Effort) Graph for Calibrated, Uncalibrated, and Actual Effort

Calibrated Model. As the MMRE for the calibrated model was approximately .049, it was highly acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .051, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 100% of the predicted values fell within 25% of their actual values. This value far exceeded the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a slight positive bias.

The three statistical tests, MMRE, RRMS, and PRED*(l)*, demonstrated the model was highly acceptable for this category.

Uncalibrated Model. As the MMRE for the uncalibrated model was approximately .049, it was highly acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .051, a value much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 100% of the predicted values fell within 25% of their actual values. This value far exceeded the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a slight positive bias.

The three statistical tests, MMRE, RRMS, and PRED*(l)*, demonstrated the model was highly acceptable for this category.

Comparison Between the Calibrated Model and the Uncalibrated Model. There was no change in the MMRE, RRMS, or the PRED*(l)* statistics. These statistics were all acceptable before being calibrated. Since the estimates were the same for both the uncalibrated and calibrated model, calibration did not improve estimating accuracy.

MIS (Schedule). This category was calibrated using function points to include schedule data. It was further stratified on the COBOL language and thirteen points were used, seven for calibration and six for validation.

The following table provides summary statistical information:

Table 12
MIS (Schedule)

Project	Schedule (Actual)	Schedule (Uncalibrated)	MRE	Wilcoxon	Schedule (Calibrated)	MRE	Wilcoxon
MIS 4	**4**	**3.68**	**0.080**	**0.32**	**3.58**	**0.105**	**0.42**
MIS 5	**8**	**13.57**	**0.696**	**-5.57**	**6.24**	**0.220**	**1.76**
MIS 6	**9**	**6.51**	**0.277**	**2.49**	**6.44**	**0.284**	**2.56**
MIS 7	**15**	**10.78**	**0.281**	**4.22**	**8.61**	**0.426**	**6.39**
MIS 8	**55**	**39.49**	**0.282**	**15.51**	**26.94**	**0.510**	**28.06**
MIS 9	**54**	**39.57**	**0.267**	**14.43**	**27.14**	**0.497**	**26.86**
		MMRE	**0.314**		**MMRE**	**0.292**	
		RMS	**9.164**		**RMS**	**14.926**	
		RRMS	**0.379**		**RRMS**	**0.721**	
		PRED(l)	**0.167**		**PRED(l)**	**0.333**	

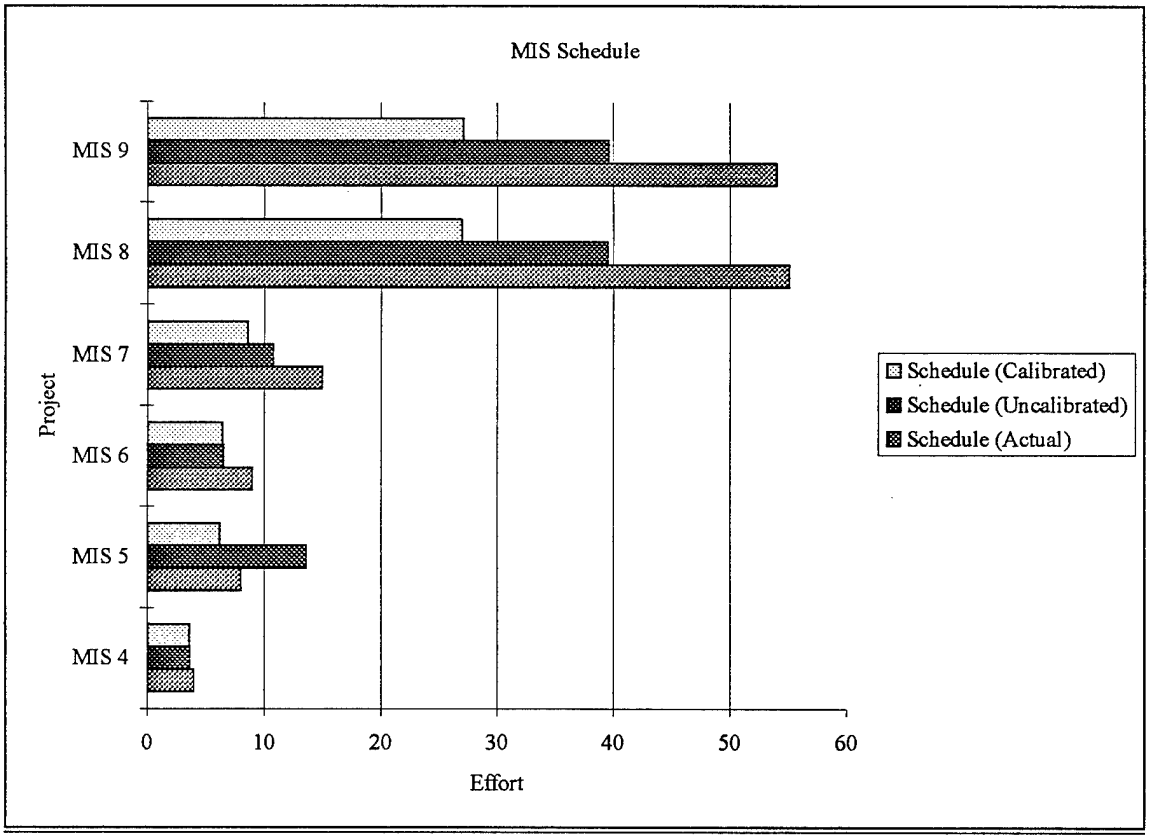

Figure 9. MIS (Schedule) Graph for Calibrated, Uncalibrated, and Actual Schedule

Calibrated Model. As the MMRE for the calibrated model was approximately .292, it was not acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also made the model unacceptable since its value was approximately .721, a value much larger than the acceptable 0.25.

When analyzing the PRED (0.25) value, approximately 34% of the predicted values fell within 25% of their actual values. This value was far less than the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a positive bias.

The three statistical tests demonstrated the model was not acceptable for this category.

Uncalibrated Model. As the MMRE for the uncalibrated model was approximately .314, it was not acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also made the model unacceptable since its value was approximately .379, a value much larger than the acceptable 0.25.

When analyzing the PRED (0.25) value, approximately 17% of the predicted values fell within 25% of their actual values. This value was far less than the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a positive bias.

The three statistical tests demonstrated the model was not acceptable for this category.

Comparison Between the Calibrated Model and the Uncalibrated Model. The MMRE decreased by approximately .022, making it 7.02% more accurate. The RRMS increased approximately .341, making it unacceptable for this category. The PRED*(l)* increased by .17. However, it was still unacceptable for this category.

Unmanned Space (Schedule). This category was calibrated using effort, SLOC, and schedule data. It was not further stratified on language and eleven points were used, six for calibration and five for validation.

The following table provides summary statistical information:

Table 13
Unmanned Space (Schedule)

Project	Schedule (Actual)	Schedule (Uncalibrated)	MRE	Wilcoxon	Schedule (Calibrated)	MRE	Wilcoxon
SigProc 7	**23**	**9.20**	**0.600**	**13.80**	**9.20**	**0.600**	**13.80**
SigProc 8	**52**	**20.80**	**0.600**	**31.20**	**20.80**	**0.600**	**31.20**
SigProc 9	**40**	**16.00**	**0.600**	**24.00**	**16.00**	**0.600**	**24.00**
SigProc 10	**40**	**16.00**	**0.600**	**24.00**	**16.00**	**0.600**	**24.00**
SigProc 11	**40**	**16.00**	**0.600**	**24.00**	**16.00**	**0.600**	**24.00**
		MMRE	**0.600**		**MMRE**	**0.500**	
		RMS	**24.049**		**RMS**	**21.954**	
		RRMS	**0.617**		**RRMS**	**0.676**	
		PRED(l)	**0.000**		**PRED(l)**	**0.000**	

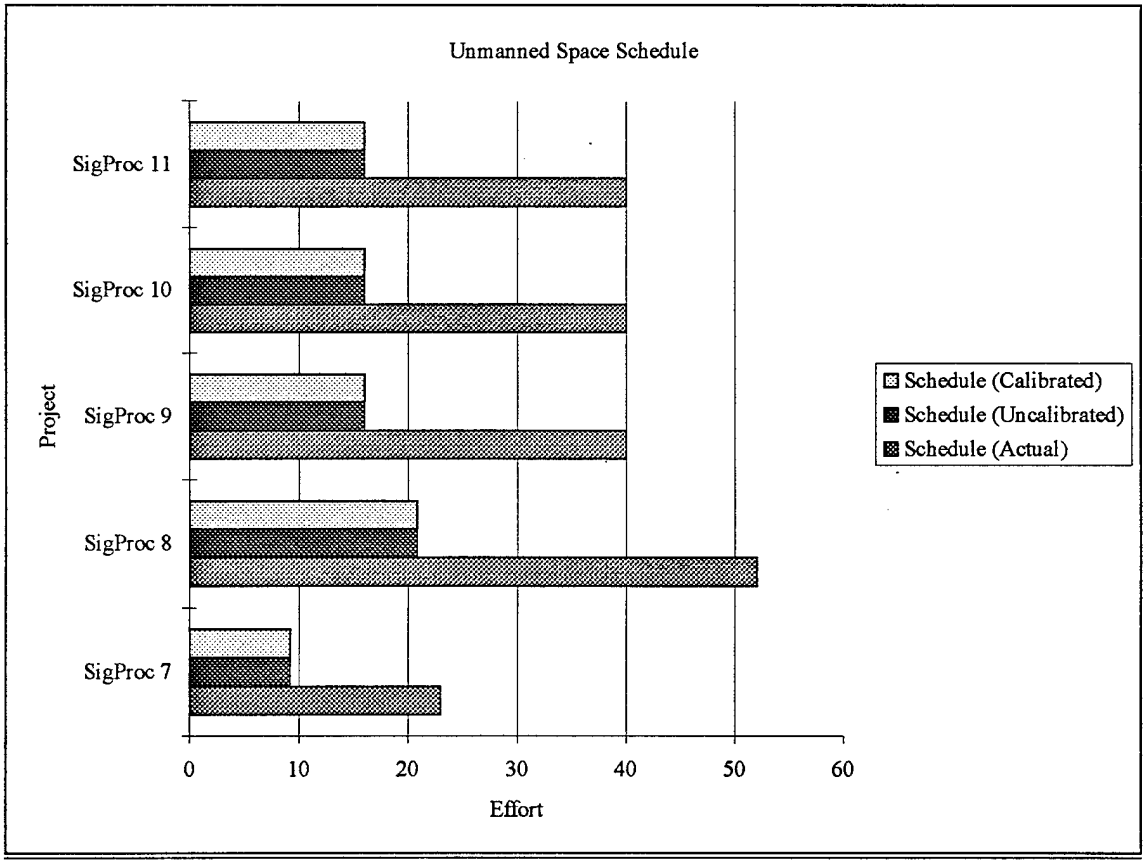

Figure 10. Unmanned Space (Schedule) Graph for Calibrated, Uncalibrated, and Actual Schedule

Calibrated Model. As the MMRE for the calibrated model was .50, it was not acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also made the model unacceptable since its value was approximately .676, a value much larger than the acceptable 0.25.

When analyzing the PRED (0.25) value, 0% of the predicted values fell within 25% of their actual values. This value was far less than the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a positive bias.

The three statistical tests demonstrated the model was not acceptable for this category.

Uncalibrated Model. As the MMRE for the uncalibrated model was .60, it was not acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also made the model unacceptable since its value was approximately .617, a value much larger than the acceptable 0.25.

When analyzing the PRED (0.25) value, 0% of the predicted values fell within 25% of their actual values. This value was far less than the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a positive bias.

The three statistical tests demonstrated the model was not acceptable for this category.

Comparison Between the Calibrated Model and the Uncalibrated Model. The MMRE decreased by approximately .1, making it 16.67% more accurate. The RRMS

increased approximately .059, making it unacceptable for this category. The PRED*(l)* did not change.

Ground in Support of Space (Schedule). This category was calibrated using effort, SLOC, and schedule data. It was further stratified on the Ada language and eight points were used, four for calibration and four for validation.

The following table provides summary statistical information:

Table 14
Ground in Support of Space (Schedule)

Project	Schedule (Actual)	Schedule (Uncalibrated)	MRE	Wilcoxon	Schedule (Calibrated)	MRE	Wilcoxon
5	**15**	**6.01**	**0.599**	**8.99**	**6.01**	**0.599**	**8.99**
6	**12**	**4.80**	**0.600**	**7.20**	**4.80**	**0.600**	**7.20**
7	**16**	**6.41**	**0.599**	**9.59**	**6.41**	**0.599**	**9.59**
8	**24**	**9.59**	**0.600**	**14.41**	**9.59**	**0.600**	**14.41**
		MMRE	**0.600**		**MMRE**	**0.600**	
		RMS	**10.396**		**RMS**	**10.396**	
		RRMS	**0.621**		**RRMS**	**0.621**	
		PRED(l)	**0.000**		**PRED(l)**	**0.000**	

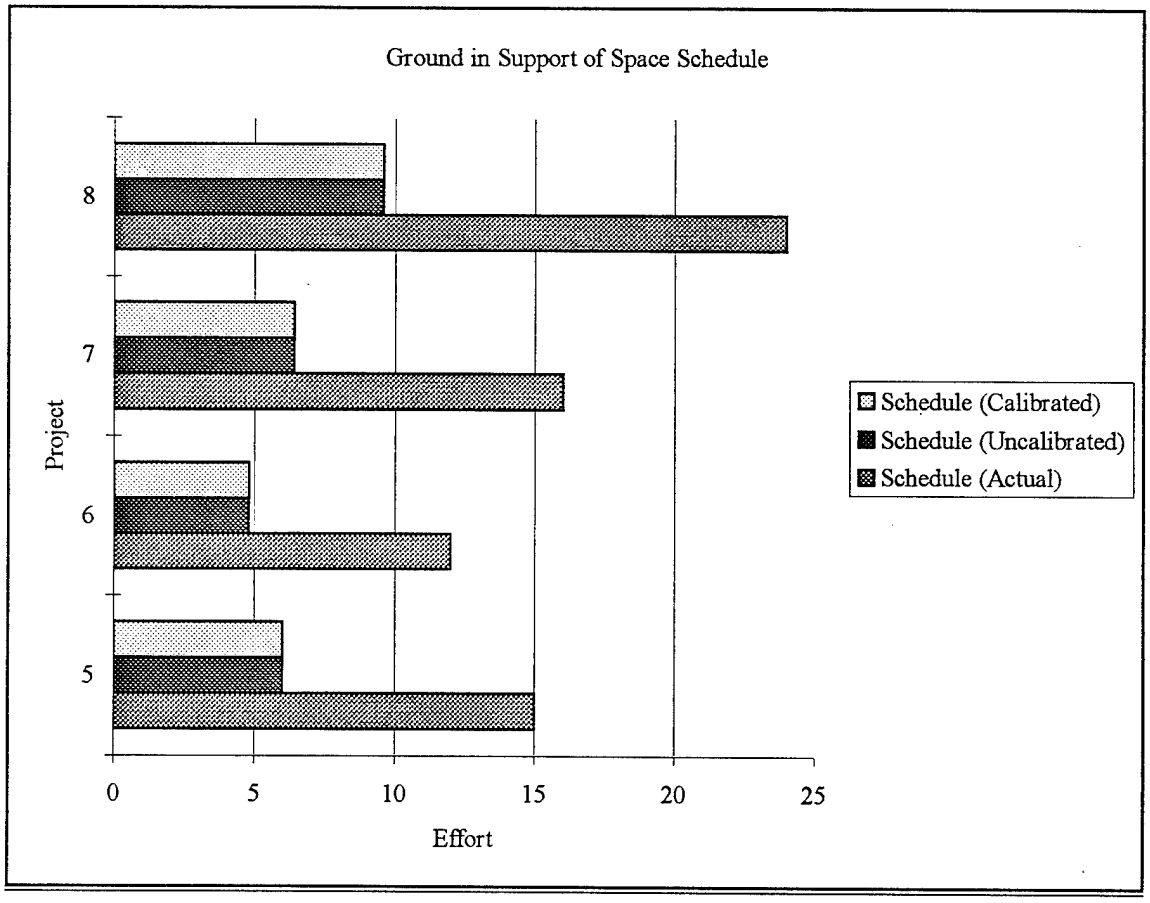

Figure 11. Ground in Support of Space (Schedule) Graph for Calibrated, Uncalibrated, and Actual Schedule

Calibrated Model. As the MMRE for the calibrated model was approximately .60, it was not acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also made the model unacceptable since its value was approximately .621, a value much larger than the acceptable 0.25.

When analyzing the PRED (0.25) value, 0% of the predicted values fell within 25% of their actual values. This value was far less than the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a positive bias.

The three statistical tests demonstrated the model was not acceptable for this category.

Uncalibrated Model. As the MMRE for the calibrated model was approximately .60, it was not acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also made the model unacceptable since its value was approximately .621, a value much larger than the acceptable 0.25.

When analyzing the PRED (0.25) value, 0% of the predicted values fell within 25% of their actual values. This value was far less than the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a positive bias.

The three statistical tests demonstrated the model was not acceptable for this category.

Comparison Between the Calibrated Model and the Uncalibrated Model. There was no change in the MMRE, RRMS, or the PRED*(l)* statistics. These statistics were all unacceptable before being calibrated.

Cobol Projects (Schedule). This category was calibrated using effort, SLOC, and schedule data. It was further stratified on the COBOL language and eight points were used, four for calibration and four for validation.

The following table provides summary statistical information:

Table 15
Cobol Projects (Schedule)

Project	Schedule (Actual)	Schedule (Uncalibrated)	MRE	Wilcoxon	Schedule (Calibrated)	MRE	Wilcoxon
5	**36**	**14.39**	**0.600**	**21.61**	**14.39**	**0.600**	**21.61**
6	**36**	**14.39**	**0.600**	**21.61**	**14.39**	**0.600**	**21.61**
7	**36**	**14.39**	**0.600**	**21.61**	**14.39**	**0.600**	**21.61**
8	**36**	**14.39**	**0.600**	**21.61**	**14.39**	**0.600**	**21.61**
		MMRE	**0.600**		**MMRE**	**0.600**	
		RMS	**21.610**		**RMS**	**21.610**	
		RRMS	**0.600**		**RRMS**	**0.600**	
		PRED(l)	**0.000**		**PRED(l)**	**0.000**	

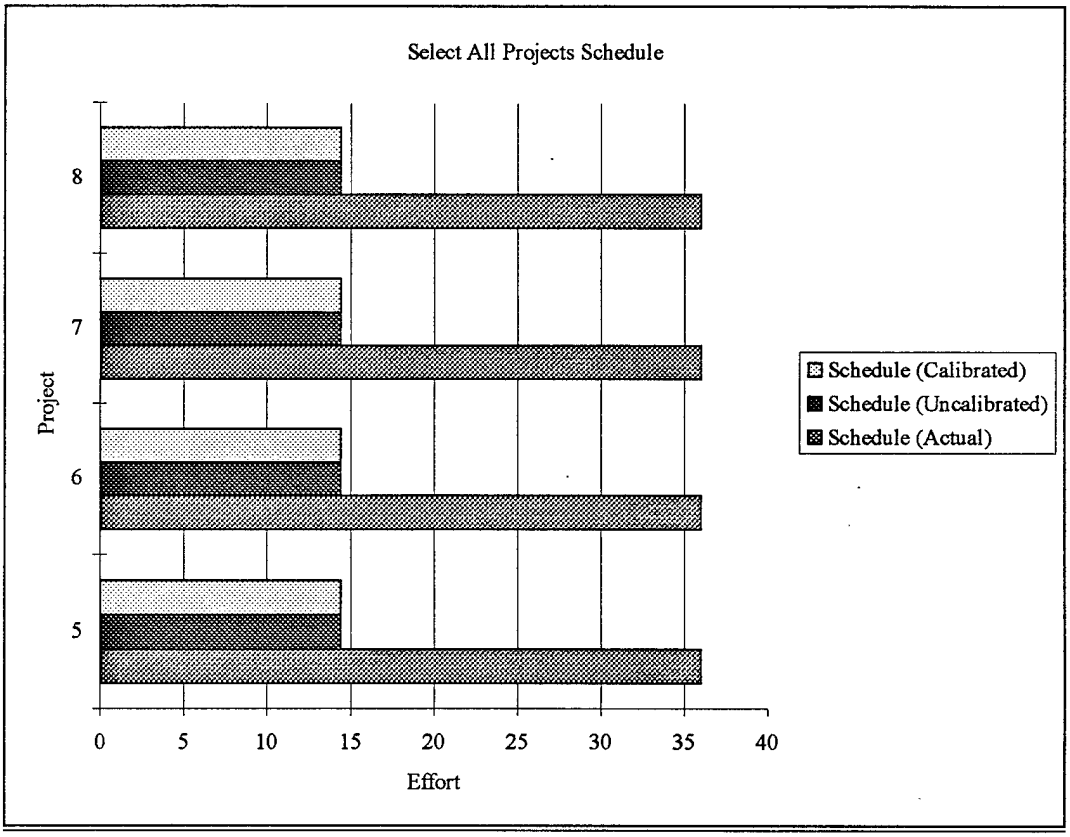

Figure 12. Cobol Projects (Schedule) Graph for Calibrated, Uncalibrated, and Actual Schedule

Calibrated Model. As the MMRE for the calibrated model was approximately .600, it was not acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also made the model unacceptable since its value was approximately .600, a value much larger than the acceptable 0.25.

When analyzing the PRED (0.25) value, 0% of the predicted values fell within 25% of their actual values. This value was far less than the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a positive bias.

The three statistical tests demonstrated the model was not acceptable for this category.

Uncalibrated Model. As the MMRE for the calibrated model was approximately .600, it was not acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also made the model unacceptable since its value was approximately .600, a value much larger than the acceptable 0.25.

When analyzing the PRED (0.25) value, 0% of the predicted values fell within 25% of their actual values. This value was far less than the 75% desired percentage to establish model accuracy.

The Wilcoxon Signed-Rank Test showed a positive bias.

The three statistical tests demonstrated the model was not acceptable for this category.

Comparison Between the Calibrated Model and the Uncalibrated Model. There was no change in the MMRE, RRMS, or the PRED*(l)* statistics. These statistics were all unacceptable before being calibrated.

Synopsis of Wilcoxon Signed-Rank Test. Four out of the eight effort categories calibrated, MIS, Military Mobile, Military - Specific Avionics, and Military Ground and Application - Command and Control showed a negative bias. The remaining four categories calibrated on effort, Military Ground and Application - Signal Processing, Unmanned Space, Ground in Support of Space, and Cobol Projects showed a positive bias.

The four categories calibrated on schedule, MIS, Unmanned Space, Ground in Support of Space, and Cobol Projects showed a positive bias.

These results make the estimating accuracy of the CHECKPOINT model even more significant because the model succeeded in all eight of the categories calibrated on effort despite calibrating on data that was not robust.

Synopsis of Calibrated Function Point Categories. Three categories, MIS, Military Mobile, and Military - Specific Avionics were calibrated using function points. The MMRE for these calibrated categories was approximately .018, .192, and .158, respectively, making the three calibrated models highly acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately. 004, .057, and .111, respectively, for the three categories. These values were much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 100%, 75%, and 75%, respectively, of the predicted values fell within 25% of their actual values. These values exceeded or equaled the 75% desired percentage to establish model accuracy.

For all three function point categories, the three statistical tests, MMRE, RRMS, and the PRED*(l)*, demonstrated the model was more than acceptable.

Synopsis of Calibrated Effort and SLOC Categories. Two categories, Military Ground and Application - Command and Control and Military Ground and Application - Signal Processing, were calibrated using effort and SLOC data points. The MMRE for these calibrated categories was approximately .165 and .090, respectively, making the two calibrated models highly acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .156 and .081, respectively, for the two categories. These values were much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 50%, and 100%, respectively, of the predicted values fell within 25% of their actual values. These values were mixed; the value for Military Ground and Application - Command & Control was less than the desired 75% and the value for Military Ground - Signal Processing was greater. With mixed results, a general statement about model accuracy for this statistic could not be made from analyzing these two categories alone.

For the two effort and SLOC categories, the MMRE and the RRMS demonstrated the model was more than acceptable. However, the PRED*(l)* value for Military Ground and Application - Command & Control did not support this hypothesis.

Synopsis of Calibrated Effort, SLOC, and Schedule Categories. Three categories, Unmanned Space, Ground in Support of Space, and Cobol Projects, were calibrated using effort, SLOC, and schedule data points. The MMRE for these calibrated categories was approximately .040, .050, and .049, respectively, making the three calibrated models highly acceptable according to Conte's criteria (Conte, Dunsmore, and Shen, 148-176).

The second statistic analyzed, the RRMS, also supported this hypothesis since its value was approximately .055, .058, and .051, respectively, for the three categories. These values were much lower than the acceptable 0.25.

When analyzing the PRED (0.25) value, 100%, 100%, and 100%, respectively, of the predicted values fell within 25% of their actual values. These values far exceeded the 75% desired percentage to establish model accuracy.

For the three effort, SLOC, and schedule categories, the three statistical tests, MMRE, RRMS, and the PRED*(l)*, demonstrated the model was highly acceptable.

Comparison of Calibrated Function Point Categories with Calibrated Effort and SLOC Categories. For the three function point categories, the three statistical tests, MMRE, RRMS, and PRED*(l)*, demonstrated the model was highly acceptable. In comparison to the two effort and SLOC categories, the MMRE and the RRMS both demonstrated the model was acceptable. The PRED*(l)* value supported this finding for

the Military Ground and Application - Signal Processing both not for the Military Ground and Application - Command and Control.

Comparison of Calibrated Function Point Categories with Calibrated Effort, SLOC, and Schedule Categories. For the three function point categories and the three effort, SLOC, and schedule categories, the three statistical tests, MMRE, RRMS, and PRED*(l)*, demonstrated the model more than was acceptable.

Synopsis of Calibrated Schedule Categories. For the four categories calibrated on schedule, MIS, Unmanned Space, Ground in Support of Space, and Cobol Projects, the three statistics, MMRE, RRMS, and PRED (0.25), demonstrated the model was not acceptable.

Contrast Between Calibrated Effort and Calibrated Schedule. The MMRE and the RRMS supported the acceptability of the model for each of the categories calibrated on function points or effort, SLOC, and schedule. When categories were calibrated on effort and SLOC only, without schedule, both the MMRE and the RRMS made the model more than acceptable. The PRED*(l)* also made the model more than acceptable in all categories except the Military Ground and Application - Command and Control . In contrast with the four categories calibrated on schedule, the three statistics demonstrated that the model was unacceptable.

Analysis of the Calibrated Data Compared with the Uncalibrated Data. For the categories calibrated on function points, MIS, Military Mobile, and Military - Specific Avionics, values for MMRE increased in accuracy by 96.71%, 86.14%, and 80.71%, respectively. The RRMS increased in accuracy by 95.91%, 86.27%, and , 83.84%

respectively. The PRED*(l)* increased by approximately .34, .5, and .25, respectively, making it a more than acceptable estimator for model accuracy.

For the models calibrated on effort and SLOC, Military Ground and Application - Command and Control and Military Ground and Application - Signal Processing, values for MMRE increased in accuracy by 14.29% for Military Ground and Application - Command and Control and did not change for Military Ground and Application - Signal Processing. The RRMS increased in value for the Command and Control variant but did not change for Signal Processing. It was acceptable in predicting model accuracy for both variants. The PRED*(l)* did not change for either category.

For the categories calibrated on effort, SLOC, and schedule, Unmanned Space, Ground in Support of Space, and Cobol Projects, the MMRE increased in accuracy by 16.67% for Unmanned Space and did not change for the other two categories. The RRMS increased in value for Unmanned Space and did not change for the other two categories; it remained an acceptable predictor for model accuracy. The PRED*(l)* did not change but it was already at 100% for the uncalibrated model.

For the categories calibrated on schedule, MIS, Unmanned Space, Ground in Support of Space, and Cobol Projects, the MMRE increased in accuracy by 7.02% for MIS and 16.67% for Unmanned Space. There was no change in this statistic for Ground in Support of Space and Cobol Projects. Although calibration did improve this value for two of the categories, it did not make this statistic an acceptable measure for predicting model accuracy for any of the categories. The RRMS was also unacceptable for the four categories. The PRED*(l)* increased by .17 for MIS. However, it was still unacceptable

for this category. The PRED*(l)* did not change for the other three categories and remained unacceptable.

The CHECKPOINT model clearly excelled in increasing cost estimation accuracy when function points were used as the inputs to the model. SLOC had far less of an impact because they are an output rather than an input. Maximum benefit from the model is derived through the use of function points. This fact comes as no surprise since the model is designed to be used with function points.

Summary

This chapter presented the analysis of the data and gave the results. The chapter described the assumptions made about missing data, adjustments made to the data, results of the calibrations for each of the eight categories, and a comparison of categories calibrated on function points with those calibrated on effort and SLOC as well as those calibrated on effort, SLOC, and schedule. A contrast between the results obtained for effort and schedule was also presented. In addition, the calibrated results for each category were compared to the uncalibrated results and an analysis of the overall improvement in the model's estimating accuracy was presented.

V. Conclusions and Recommendations for Follow-on Research

Overview

This chapter provides concluding and summary comments for each of the eight calibrated templates:

- MIS
- Military Mobile
- Military -- Specific Avionics
- Military Ground an Application -- Command & Control
- Military Ground and Application - Signal Processing
- Unmanned Space
- Ground in Support of Space
- Cobol Projects

The objective of this research effort was twofold: first, to aid DoD decision makers by providing a calibration method, based on the most current data available, that may improve CHECKPOINT's accuracy in predicting future project software effort (cost) and second, to provide the reader with a step-by-step reference on how to calibrate the current version of CHECKPOINT to their own database.

Now that CHECKPOINT's calibration procedure has been identified in writing, it can be incorporated in the next version of the User's Guide. This will standardize the procedures possibly used by independent corporations and encourage others that have not yet attempted to calibrate the model to conduct the procedure using their own data.

Limitations

One limitation of this study was the SWDB was not originally designed for use in calibrating the CHECKPOINT model. As such, only three categories; MIS, Military Mobile, and Military - Specific Avionics, could be calibrated using function points. The other five categories; Military Ground and Application - Command and Control, Military Ground and Application - Signal Processing, Unmanned Space, Ground in Support of Space, and Cobol Projects, were calibrated using the values for effort, size, and schedule.

A second limitation of this study was only four categories; MIS, Military Mobile, Ground in Support of Space, and Cobol Projects could be further stratified into language. The other four categories contained a combination of languages, making their calibration factors less accurate.

Summary of Results

The following table provides MMRE, RRMS, and PRED*(l)* summary statistics for the categories both before and after calibration. The before or uncalibrated statistics are denoted by U and after or calibrated statistics are denoted by C. The eight categories calibrated on effort are denoted with (E) and the four categories calibrated on schedule are denoted with (S):

Table 16
Summary Statistics for the Eight Categories

Category	MMRE-U	RRMS-U	PRED(l)-U	MMRE-C	RRMS-C	PRED(l)-C
MIS (E)	**0.542**	**0.101**	**0.667**	**0.018**	**0.004**	**1.000**
Military Mobile (E)	**1.384**	**0.412**	**0.250**	**0.192**	**0.057**	**0.750**
Military Specific Avionics (E)	**0.817**	**0.685**	**0.500**	**0.158**	**0.111**	**0.750**
Military Ground and App. - C&C (E)	**0.193**	**0.145**	**0.500**	**0.165**	**0.156**	**0.500**
Military Ground and App. - SP (E)	**0.090**	**0.081**	**1.000**	**0.090**	**0.081**	**1.000**
Unmanned Space (E)	**0.048**	**0.050**	**1.000**	**0.040**	**0.055**	**1.000**
Ground in Support of Space (E)	**0.050**	**0.058**	**1.000**	**0.050**	**0.058**	**1.000**
Cobol Projects (E)	**0.049**	**0.051**	**1.000**	**0.049**	**0.051**	**1.000**
MIS (S)	**0.314**	**0.379**	**0.167**	**0.292**	**0.721**	**0.333**
Unmanned Space (S)	**0.600**	**0.617**	**0.000**	**0.500**	**0.676**	**0.000**
Ground in Support of Space (S)	**0.600**	**0.621**	**0.000**	**0.600**	**0.621**	**0.000**
Cobol Projects (S)	**0.600**	**0.600**	**0.000**	**0.600**	**0.600**	**0.000**

Three highly significant results were obtained in conducting this effort:

1. CHECKPOINT, according to Conte's criteria, successfully calibrated seven out of the eight categories on effort. The PRED*(l)* value for Military Ground and Application - Command and Control was the only statistic that was not satisfied.
2. CHECKPOINT unsuccessfully calibrated four (out of four) categories on schedule.
3. Calibration increased accuracy by as much as 96.71%.

Recommendations for Follow-on Research

This study could be repeated calibrating on schedule instead of effort for the five models that comprise *The Pentateuch Study*. Another model, SAGE, could be calibrated on effort, schedule, or both. In addition, models calibrated to date which include PRICE-S, REVIC, SASET, SEER-SEM, SLIM, and SOFTCOST-OO could be stratified further by language.

Of specific interest to one local area contractor are:

1. Calibrating on additional data bases that could be used to validate the results obtained using the SWDB.
2. Conducting calibrations using CASE technology, i.e. IEF and NEXT.
3. Calibrating on the Ada language only.
4. Conducting more calibrations using only function points.

Appendix A: Glossary

The following are some useful definitions for understanding the results of this research:

Accuracy - "The degree of error in an estimate. An estimate is said to be more accurate if the amount of error in that estimate is reduced" (Kressin, 1995:73).

Algorithm - "A mathematical set or ordered steps leading to the optimal solution of a problem in a finite number of operations" (Coggins and Russell, 1993:5).

Calibration - "adjustment of the model equations to induce the model to provide a predicted outcome as close as possible to the actual outcome for a given set of data" (Vegas, 1995:5).

CHECKPOINT - software cost estimating model developed by Capers Jones and distributed by SPR (Ferens, 1995).

COCOMO - "The Constructive Cost Model, a software cost estimating model developed by Barry Boehm" (Weber, 1995:A-1).

Cost Estimating - "The art of collecting and scientifically studying costs and related information on current and past activities as a basis for projecting costs as an input to the decision making process for a future activity" (Coggins and Russell, 1993:5).

Cost Model - "A tool consisting of one or more cost estimating relationships, estimating methodologies, or estimating techniques and used to predict the cost of a system or its components" (Coggins and Russell, 1993:5).

CSCI, CSC, and CSU - "Large software development efforts are generally broken down into smaller, more manageable entities called computer software configuration items (CSCIs). Each CSCI may be further broken down into computer system components (CSCs) and each CSC may be further broken down into computer software units (CSUs)" (Weber, 1995:A-1).

Effort - "the number of person hours or person months required to produce function points or SLOC" (Vegas, 1995:7).

Function Point - a concept for computing software size from five attributes: external inputs, external outputs, external inquiries, external interfaces, and internal files (Software Productivity Research, 1995:1-3).

IFPUG - International Function Point User's Group, an organization devoted to continuous research and update of function points (Ferens, 1995).

Magnitude of Relative Error (MRE) - "A measure of accuracy that reflects the degree of error in a particular estimate. Specifically, it is the absolute difference between an estimate and actual observation, divided by the actual observation" (Conte, Dunsmore, and Shen, 1986:172).

MCR - Management Consulting & Research, Inc. A DoD support contractor responsible for the design, development, implementation, and maintenance of the SMC SWDB (Stukes, 1995).

Mean Magnitude of Relative Error (MMRE) - "A measure of accuracy that reflects the average degree of error produced by a set of estimates. Specifically, it is the sum of the individual MRE measures for a set of estimates divided by the number of estimates" (Conte, Dunsmore, and Shen, 1986:172).

Normalization - "The process of rendering constant or adjusting for known differences" (Weber, 1995:A-2).

Person months - "a measurement unit of the effort required to produce a software program; the standard is 152 hours of labor per person month" (Vegas, 1995:8).

Prediction at Level k/n (PRED(k/n) - "Sometimes referred to as the percentage method, the PRED(*k/n*) is a measure that represents the percentage of estimates that fall within a predefined amount of their actuals. Specifically, it is the percentage of estimates in a set of estimates whose MRE is less than or equal to a preset value (*k/n)"* (Conte, Dunsmore, and Shen, 1986:172).

PRICE-S - "Programmed Review of Information for Costing and Evaluation-Software;" model developed with the combined experience and input of government and commercial software developers. "The model was specifically created to assist project managers in assessing values for cost, time, and manpower based on the historical data of previous projects" (Galonsky, 1995:1-2).

Regression Analysis - describes the relationship between at least two variables, one dependent and one or more independent variables.

REVIC - "A software cost estimating model developed by Raymond Kile" (Weber, 1995:A-2).

SASET - "'Software Architecture Sizing and Estimating Tool;' a parametric software cost estimating model developed by Lockheed Martin" (Vegas, 1995:8).

SEER-SEM - "'System Evaluation and Estimation of Resources Software Estimation Model' developed by Galorath Associates, Inc." (Rathmann, 1995:15).

SLIM - "'Software Lifecycle Model' developed by Larry Putnam and distributed through Quantitative Software Management, Incorporated" (Kressin, 1995:20).

SLOC - "source lines of code" is a measurement unit of the size of a software program. "In this study, logical lines of code are used, which means an instruction may take up more than one physical line of code or two or more instructions could be on a single line" (Stukes, 1996). This counting convention does not include blank lines, comments, unmodified vendor supplied operating system or utility software, or other non-developed code. It includes only executable program instructions created by the project personnel which are delivered in the final product (Rathmann, 1995:7).

SMC - "Air Force Space and Missile Systems Center, responsible for research, development, acquisition, on-orbit testing, and sustainment of military space and missile systems" (Tinkler, 1996).

SMC/FMC - SMC Directorate of Cost; co-developer of the SMC database (Tinkler, 1996).

Software - The combination of computer programs, data, and documentation which enables computer equipment to perform computational or central functions (Weber, 1995:A-3).

Software Development Cycle - The software development cycle is typically broken into eight phases: (1) Systems Requirements Analysis and Design, (2) Software Requirements Analysis, (3) Preliminary Design, (4) Detailed Design, (5) Code and CSU Testing, (6) CSC Integration and Testing, (7) CSCI Testing, and (8) System Testing (Ferens, 1995).

SPR - Software Productivity Research, Incorporated, developer of the Checkpoint software estimating model (Software Productivity Research, 1993:1-3).

Stratification - "the division of data into homogenous groups in order to perform analysis and discover patterns; more detailed subdivisions usually reduce the number of useable points" (Vegas, 1995:8).

SWDB - "Software Database" is the database created by SMC and used for this research effort. Version 2.1 is the version used to support this study (Stukes, 1996).

Validation - "process of determining the accuracy of the model; the difference between the model's predicted outcome and the actual outcome for a set of data similar, but not identical, to the set used in calibration" (Vegas, 1995:9).

Wilcoxon Signed-Ranked Test - "A non-parametric test that indicates whether there is bias in a set of observations" (Mendenhall, Wackerly, and Scheaffer, 1990:680).

Appendix B: Data Records

MIS Function Points

PRODUCT NUMBER	[IT] INPUTS	[OT] OUTPUTS	[QT] QUERIES	[FT] FILES	[ET] EXTERNALS	LOC	MTHS	PERSON MTHS
MIS 1	?	?	?	?	?	69,866	?	?
MIS 2	?	?	?	?	?	95,807	?	?
MIS 3	?	?	?	?	?	261,619	?	42
	3.25	7	5	4	5	?	9	?
MIS 4	?	?	?	?	?	92,701	?	?
MIS 5	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 6	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 7	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 8	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 9	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 10	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 11	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 12	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 13	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 14	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 15	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 16	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 18	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 19	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 20	?	?	?	?	?	30,523	?	36
MIS 21	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 22	8	4	4	7	22	N/A	N/A	N/A
	8	**3**	**4**	**7**	**23**	**22,488**	**13**	**79**
MIS 23	**30**	**67**	**23**	**28**	**50**	**129,412**	**?**	**203**
MIS 24	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 25	?	?	?	?	?	87,300	?	13
MIS 27	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	?	?	?	?	?	302,204	?	302
MIS 28	?	?	?	?	?	68,119	?	29
MIS 29	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 30	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 31	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 32	2	5.6	5	7	14	N/A	N/A	N/A
	0	5.2	0	2	11	?	?	?
MIS 33	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 34	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 36	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 37	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 38	**25**	**43.4**	**41**	**30**	**45**	**231,833**	**?**	**451**
MIS 39	17	30.2	23	25	29	N/A	N/A	N/A
	15	55.4	34	25	27	N/A	N/A	N/A
	27	68.4	44	25	62	N/A	N/A	N/A
MIS 40	10	10.4	17	25	47	N/A	N/A	N/A
	12	**13**	**10**	**25**	**35**	**264,184**	**24**	**354**
MIS 42	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 43	20	57	11	20	67	N/A	N/A	N/A
	21	60	12	28	82	N/A	N/A	N/A
	22	**60.6**	**12**	**28**	**84**	**126,455**	**30**	**335**

MIS Function Points

PRODUCT NUMBER	[IT] INPUTS	[OT] OUTPUTS	[QT] QUERIES	[FT] FILES	[ET] EXTERNALS	LOC	MTHS	PERSON MTHS
MIS 44	5	4.8	6	37	38	9,524	4	10
MIS 45	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	29	31	39	10	6	4,408	8	6
MIS 46	14.25	9.8	18	14	14	17,082	9	30
MIS 47	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	47	18.6	101	37	23	N/A	N/A	N/A
	29	43.8	42	18	27	N/A	N/A	N/A
	31.5	25.2	34	18	21	88360E	?	114
MIS 48	18	28	42	16	38	N/A	N/A	N/A
	20	10	46	16	38	80989E	?	247
MIS 49	8	29	97	13	53	N/A	N/A	N/A
	17	146	85	13	53	179377E	?	156
MIS 50	?	?	?	?	?	145300E	?	?
MIS 53	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 54	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 55	0	9.2	0	8	36	N/A	N/A	N/A
	6.5	7.8	7	17	22	N/A	N/A	N/A
	2	0.6	2	3	25	56,236	15	132
MIS 56	1.5	6	1	3	18	N/A	N/A	N/A
	1.5	6	1	3	18	19860E	?	?
MIS 57	1	19	1	16	27	N/A	N/A	N/A
	1	31	1	16	28	128,172	55	296
MIS 58	43	98	113	60	49	N/A	N/A	N/A
	42	124	143	60	52	N/A	N/A	N/A
	20.75	30.6	49	30	98	234,516	54	673
MIS 59	21	117	0	107	60	N/A	N/A	N/A
	14	91	39	103	0	N/A	N/A	N/A
	58	201	77	4	70	N/A	N/A	N/A
	18	128.4	88	46	83	N/A	N/A	N/A
	16	110.4	67	53	98	299,543	54	753
MIS 60	25	18	34	60	33	N/A	N/A	N/A
	23.25	13.2	31	60	49	73,016	27	276
MIS 61.1	33.25	79.2	45	63	42	N/A	N/A	N/A
	38.25	83.2	40	57	98	N/A	N/A	N/A
	9.25	6.2	8	8	34	N/A	N/A	N/A
	47.5	89.4	48	60	104	147,346	28	224
MIS 61.2	66.25	95.2	52	44	83	?	?	?
MIS 62	52.25	33.8	67	43	58	N/A	N/A	N/A
	50.25	33.8	65	43	58	?	?	?
MIS 63	?	?	?	?	?	?	?	?
	?	?	?	?	?	?	?	?
	5.25	12	6	13	14	?	?	?
MIS 64	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 65	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 66	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 68	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 69	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MIS 70	5.25	36.8	15	1	24	83,028	20	101
x								

Military Mobile

Function Points

Number of records included in search: 15

Application	Ext Inpts	Ext Outpts	Ext Inq	Int Files	Ext Int
Database	11	25			
Signal Processing			0	10	5
Other (MM1)	8			8	7
Command/Control (MM2)	35	23	9		
Other (MM3)	15	15			
Mission Planning					
Database	7	2		25	
Signal Processing	2	2	1	4	2
Command/Control (MM4)	304	304	304	3	0
MMI/Graphics (MM5)	300	300	300	25	0
Command/Control (MM6)	300	300	300	1	0
Command/Control (MM7)	300	300	300	1	0
Other					
Command/Control					
Command/Control (MM8)	120	95		85	

Military Mobile

Effort, SLOC, and Sch

Number of records included in search: 15

Application	Tot Eff	Norm Eff Sz	Sch Mon	Prog Lang
Database	88.00	17134		Assembly C
Signal Processing	250.00	30000		Assembly 50% PASCAL 50%
Other (MM1)	41.00	2311		Ada 95% Machine 5%
Command/Control (MM2)	418.00	18052		Ada 90% C 9% Machine 1%
Other (MM3)	59.00	3268		Ada 95% Machine 5%
Mission Planning	233.00	63254	30	Ada 8% FORTRAN 92%
Database	300.00	697814	60	Ada 20% C 30% FORTRAN 50%
Signal Processing	10.50	1958	12	Assembly 25% C 75%
Command/Control (MM4)	743.60	26239	59	Ada 95% Assembly 5%
MMI/Graphics (MM5)	920.00	32464	59	Ada 95% Assembly 5%
Command/Control (MM6)	211.10	7448	59	Ada 95% Assembly 5%
Command/Control (MM7)	179.00	6317	59	Ada 95% Assembly 5%
Other	759.90	26814	59	Ada 95% Assembly 5%
Command/Control	1666.10	58789	59	Ada 95% Assembly 5%
Command/Control (MM8)	15.00	15025	15	Ada 100%

Military - Specific Avionics

Function Points

Number of records included in search: 12

Application	Ext Inpts	Ext Outpts	Ext Inq	Int Files	Ext Int
Command/Control MMI/Graphics MMI/Graphics Process Control Command/Control (Milspav1)	294	358	1	6	
Signal Processing (Milspav2)	20	28	3	12	4
Diagnostics (Milspav3)				200	
Command/Control (Milspav4)	40	50	100	20	5
Command/Control (Milspav5)	250	875	50	40	0
Command/Control (Milspav6)	120	117	0	139	4
Simulation (Milspav7)	2	2	0	0	1
OS/Executive (Milspav8)	6	6	0	2	124

Military - Specific Avionics

Effort, SLOC, and Sch

Number of records included in search: 12

Application	Tot Eff	Norm Eff Sz	Sch Mon	Prog Lang
Command/Control (Milspav1)	390.00	43207		JOVIAL 95%
MMI/Graphics (Milspav2)	209.00	32878		JOVIAL 95%
MMI/Graphics (Milspav3)	118.00	22027		JOVIAL
Process Control (Milspav4)	793.00	58153		JOVIAL
Command/Control (Milspav5)	490.00	22148		JOVIAL 85%
Signal Processing (Milspav6)	54.00	4144		JOVIAL 100%
Diagnostics (Milspav7)	400.00	45353		Assembly JOVIAL
Command/Control (Milspav8)	690.00	40000		Other
Command/Control	259.00	33158	25	Ada 98% Assembly 2%
Command/Control	37.00	37000	18	Assembly 1% C 1% FORTRAN 98%
Simulation	68.00	18000	30	Ada 100%
OS/Executive	409.00	26000	44	Ada 99% Assembly 1%

Mil Ground - Command & Control

Effort, SLOC, and Sch

Number of records included in search: 13

Application	Tot Eff	Norm Eff Sz	Sch Mon	Prog Lang
Command/Control (MGCC1)	10.00	1500		
Command/Control (MGCC2)	127.00	45057		Assembly 35% FORTRAN 65%
Command/Control (MGCC3)	545.00	128200		Assembly 8% FORTRAN 92%
Command/Control (MGCC4)	656.00	144000		FORTRAN
Command/Control (MGCC5)	95.00	25842		
Command/Control (MGCC6)	139.00	23881		
Command/Control (MGCC7)	322.00	162039		
Command/Control (MGCC8)	101.00	18560		
Command/Control (MGCC9)	100.00	21681		
Command/Control (MGCC10)	286.00	69772		
Command/Control (MGCC11)	74.00	8398		
Command/Control (MGCC12)	181.20	43437	0	C 100%
Command/Control (MGCC13)	196.00	85214	48	Assembly 50% C 25% FORTRAN 25%

Mil Ground - Command & Control

Function Points

Number of records included in search: 13

Application	Ext Inpts	Ext Outpts	Ext Inq	Int Files	Ext Int
Command/Control (MGCC1)					
Command/Control (MGCC2)					
Command/Control (MGCC3)					
Command/Control (MGCC4)					
Command/Control (MGCC5)					
Command/Control (MGCC6)					
Command/Control (MGCC7)					
Command/Control (MGCC8)					
Command/Control (MGCC9)					
Command/Control (MGCC10)					
Command/Control (MGCC11)					
Command/Control (MGCC12)					
Command/Control (MGCC13)					

Mil Ground - Signal Processing

Effort, SLOC, and Sch

Number of records included in search: 20

Application	Tot Eff	Norm Eff Sz	Sch Mon	Prog Lang
Signal Processing (Milgrs1)	134.00	45035		Assembly 7% PASCAL 93%
Signal Processing (Milgrs2)	165.00	47965		
Signal Processing (Milgrs3)	13.00	16016		
Signal Processing (Milgrs4)	738.00	71851		
Signal Processing (Milgrs5)	192.00	29147		
Signal Processing (Milgrs6)	278.00	46595		
Signal Processing (Milgrs7)	645.00	123710		
Signal Processing (Milgrs8)	228.00	44527		
Signal Processing (Milgrs9)	264.00	23787		
Signal Processing (Milgrs10)	154.00	12121		
Signal Processing (Milgrs11)	274.00	60233		
Signal Processing (Milgrs12)	190.00	14389		
Signal Processing (Milgrs13)	6.00	70020		
Signal Processing (Milgrs14)	348.00	28782		
Signal Processing (Milgrs15)	86.00	23703		
Signal Processing (Milgrs16)	145.00	29802		
Signal Processing (Milgrs17)	192.00	31720		
Signal Processing (Milgrs18)	149.00	11534		
Signal Processing (Milgrs19)	109.00	8965		
Signal Processing (Milgrs20)	113.00	50000	18	Ada 100%

Mil Ground - Signal Processing

Function Points

Number of records included in search: 20

Application	Ext Inpts	Ext Outpts	Ext Inq	Int Files	Ext Int
Signal Processing (Milgrs1)	2	2			1
Signal Processing (Milgrs2)					
Signal Processing (Milgrs3)					
Signal Processing (Milgrs4)					
Signal Processing (Milgrs5)					
Signal Processing (Milgrs6)					
Signal Processing (Milgrs7)					
Signal Processing (Milgrs8)					
Signal Processing (Milgrs9)					
Signal Processing (Milgrs10)					
Signal Processing (Milgrs11)					
Signal Processing (Milgrs12)					
Signal Processing (Milgrs13)					
Signal Processing (Milgrs14)					
Signal Processing (Milgrs15)					
Signal Processing (Milgrs16)					
Signal Processing (Milgrs17)					
Signal Processing (Milgrs18)					
Signal Processing (Milgrs19)					
Signal Processing (Milgrs20)					

Unmanned Space

Effort, SLOC, and Sch

Number of records included in search: 39

Application	Tot Eff	Norm Eff Sz	Sch Mon	Prog Lang
Command/Control	615.00	80000		
OS/Executive	52.00	2000		Ada 95% Assembly 5%
Command/Control	798.00	6000		
Command/Control	204.00	1950		
Signal Processing	200.00	6000		
Command/Control	53.00	600		
Signal Processing	106.00	600		
Mission Planning	30.00			
Command/Control	125.00			
Command/Control	7.00	600		
Signal Processing	191.00	600		
Command/Control	1511.00	8290		
Command/Control	1248.00	19500		
OS/Executive	145.00	12810		Ada 30% C 70%
OS/Executive	90.00	9334		C 100%
Command/Control	548.00		36	Other 100%
Signal Processing	51.00	5000		Other 100%
Signal Processing	140.00	13000		Assembly 100%
Signal Processing	394.00			PASCAL 100%
Signal Processing	271.00			Ada 100%
Signal Processing (SigProc1)	66.00	14000	24	C 100%
Signal Processing (SigProc2)	39.00	3000	18	Ada 100%
Signal Processing (SigProc3)	28.40	12000	20	Ada 100%
Signal Processing (SigProc4)	55.40	4000	22	Ada 100%
Signal Processing (SigProc5)	194.00	34000	15	C 100%
Signal Processing (SigProc6)	26.60	9000	8	C 100%
Signal Processing (SigProc7)	202.00	11000	23	Assembly 100%
Signal Processing	550.00	22000		Assembly 50% FORTRAN 50%
Signal Processing	63.00	5000		Assembly 50% C 50%
Signal Processing	410.00	32000	52	Assembly 100%
Signal Processing (SigProc8)	92.70	7000		C 100%
Signal Processing	764.00	30000		Assembly 100%
Signal Processing	313.00	15000		Assembly 100%
Simulation	45.00	14000		C 100%
Other	37.00	5000		COBOL 100%
Command/Control	628.60	19810		Ada 73% Assembly 27%
Command/Control	321.70	16759	40	Ada 100%
Command/Control (SigProc9)	321.70	16759	40	Ada 100%
Command/Control (SigProc10)	321.70	16759	40	Ada 100%
Command/Control (SigProc11)				

Unmanned Space

Function Points

Number of records included in search: 39

Application	Ext Inpts	Ext Outpts	Ext Inq	Int Files	Ext Int
Command/Control					
OS/Executive	5	5	10	1	3
Command/Control					
Command/Control					
Signal Processing					
Command/Control					
Signal Processing					
Mission Planning					
Command/Control					
Command/Control					
Signal Processing					
Command/Control					
Command/Control					
OS/Executive					
OS/Executive	16	30	1	24	
Command/Control					
Signal Processing					
Signal Processing					
Signal Processing					
Signal Processing					
Signal Processing (SigProc1)					
Signal Processing (SigProc2)					
Signal Processing (SigProc3)					
Signal Processing (SigProc4)					
Signal Processing (SigProc5)					
Signal Processing (SigProc6)					
Signal Processing (SigProc7)					
Signal Processing					
Signal Processing					
Signal Processing (SigProc8)					
Signal Processing					
Signal Processing					
Signal Processing					
Simulation					
Other					
Command/Control					
Command/Control (SigProc9)					
Command/Control (SigProc10)					
Command/Control (SigProc11)					

Ground in Support of Space

Effort, SLOC, and Sch

Number of records included in search: 85

Application	Tot Eff	Norm Eff Sz	Sch Mon	Prog Lang
Command/Control	64.00	4613		Assembly 100%
Command/Control	80.00	11700		JOVIAL 100%
Command/Control	912.00	116800		JOVIAL 100%
Command/Control	115.00	14000		JOVIAL 100%
Command/Control	523.00	56200		JOVIAL 100%
Command/Control	478.00	48300		JOVIAL 100%
Command/Control	432.00	50300		JOVIAL 100%
Command/Control	296.00	69450		FORTRAN 45% JOVIAL 55%
Command/Control	164.00	22900		JOVIAL 100%
Command/Control	140.00	16300		JOVIAL 100%
Command/Control	57.00	6800		JOVIAL 100%
Database	244.00	117000		
Mission Planning	602.00	225000		
Signal Processing	1055.00	95000		
Signal Processing	1169.00	52275		
Mission Planning	75.00	2920		
Command/Control	401.00	250000		
Database	530.00	80000		
Mission Planning	86.00	90300		
Signal Processing	234.00	8000		
Database	5.00	21000		
Mission Planning	206.00	16300		
Signal Processing	160.00	8000		
Database	235.00	162945		
Mission Planning	109.00	13000		
Mission Planning	1468.00	399635		
Signal Processing	652.00	66843		
Signal Processing	765.00	358000		
Command/Control	787.00	278488		
Command/Control	60.00	34650		Assembly 50% FORTRAN 50%
Command/Control	19.00	7000		C 100%
Mission Planning	74.00	60087		
Command/Control	90.00	45000		Ada 100%
Command/Control	345.00	130000		FORTRAN 100%
Command/Control	244.00	126000		FORTRAN 100%
Command/Control	18.10	16000		Assembly 100%
Command/Control	10.00	6000		Ada 100%
Command/Control	636.00	22000	60	Other 100%

Ground in Support of Space

Effort, SLOC, and Sch

Number of records included in search: 85

Application	Tot Eff	Norm Eff Sz	Sch Mon	Prog Lang
Command/Control	793.00	84000	48	Assembly 100%
Command/Control (8)	74.30	18000	24	Assembly 100%
Command/Control (1)	63.20	6000	37	Ada 100%
Command/Control	105.00	11000	24	Other 100%
Command/Control (2)	118.00	22000	14	Ada 100%
Command/Control	47.40	19000	11	FORTRAN 100%
Command/Control (3)	85.40	42000	12	Ada 100%
Command/Control	100.00	100000	20	PASCAL 100%
Command/Control	250.00	150000		Ada 100%
Command/Control	48.70	21000		Other 100%
Mission Planning	88.90	24000		PASCAL 100%
Mission Planning	50.00	19000	33	FORTRAN 100%
Mission Planning	32.00	12000	19	FORTRAN 100%
Mission Planning	70.00	35000	20	PASCAL 100%
Mission Planning (4)	35.00	24000	26	Ada 100%
Mission Planning	103.00	83000		Ada 100%
Mission Planning	12.00	11000		Ada 100%
Mission Planning (5)	15.00	11000	15	Ada 100%
Message Switching	292.00	55000		Ada 100%
Message Switching	31.00	2000		Ada 100%
Message Switching	145.00	18000	18	Other 100%
Message Switching	331.00	47000	36	Other 100%
Message Switching	234.00	29000	25	Assembly 100%
Message Switching	196.00	17000		Assembly 100%
Message Switching	278.00	50000		Assembly 100%
Signal Processing	497.00	62000		Other 100%
Signal Processing	12.00	7000		Ada 100%
Signal Processing	22.60	14000		C 50% PASCAL 50%
Signal Processing	210.00	100000	66	PASCAL 100%
Signal Processing	72.00	32000		Other 100%
Signal Processing	128.00	35000	25	PASCAL 100%
Signal Processing	140.00	10000	12	Other 100%
Signal Processing	59.00	16000		FORTRAN 100%
Signal Processing	42.00	10000	33	FORTRAN 100%
Signal Processing	120.00	45000	32	FORTRAN 100%
Signal Processing	57.70	14000	29	Other 100%
Signal Processing	221.00	40000		Other 100%
Simulation	130.00	75000	41	C 100%
Simulation	526.00	49000	33	Other 100%

Ground in Support of Space

Effort, SLOC, and Sch

Number of records included in search: 85

Application	Tot Eff	Norm Eff Sz	Sch Mon	Prog Lang
Simulation (6)	20.10	3000	12	Ada 100%
Simulation	222.00	80000		Ada 100%
Simulation	138.00	50000		FORTRAN 100%
S/W Development Tools	225.00	55000		C 100%
S/W Development Tools	36.00	12000		Ada 100%
S/W Development Tools	94.00			Ada 100%
Other (7)	71.00	55000	16	Ada 100%
Other	60.00	30000		Ada 100%

Ground in Support of Space

Function Points

Number of records included in search: 85

Application	Ext Inpts	Ext Outpts	Ext Inq	Int Files	Ext Int
Command/Control	16	9	3	1	4
Command/Control					
Command/Control					
Command/Control					
Command/Control					
Command/Control					
Command/Control					
Command/Control					
Command/Control					
Command/Control					
Command/Control					
Database					
Mission Planning					
Signal Processing					
Signal Processing					
Mission Planning					
Command/Control					
Database					
Mission Planning					
Signal Processing					
Database					
Mission Planning					
Signal Processing					
Database					
Mission Planning					
Mission Planning					
Signal Processing					
Signal Processing					
Command/Control					
Command/Control					
Command/Control					
Mission Planning					
Command/Control					
Command/Control					
Command/Control					
Command/Control					
Command/Control					
Command/Control					
Command/Control					

Ground in Support of Space

Function Points

Number of records included in search: 85

Application	Ext Inpts	Ext Outpts	Ext Inq	Int Files	Ext Int
Command/Control (8)					
Command/Control (1)					
Command/Control					
Command/Control (2)					
Command/Control					
Command/Control (3)					
Command/Control					
Command/Control					
Command/Control					
Mission Planning		1			
Mission Planning					
Mission Planning					
Mission Planning					
Mission Planning (4)					
Mission Planning					
Mission Planning					
Mission Planning (5)					
Message Switching					
Message Switching					
Message Switching					
Message Switching					
Message Switching					
Message Switching					
Message Switching					
Signal Processing					
Signal Processing					
Signal Processing					
Signal Processing					
Signal Processing					
Signal Processing					
Signal Processing					-10
Signal Processing					
Signal Processing					
Signal Processing					
Signal Processing					
Signal Processing					
Simulation					
Simulation					
Simulation (6)					

Ground in Support of Space

Function Points

Number of records included in search: 85

Application	Ext Inpts	Ext Outpts	Ext Inq	Int Files	Ext Int
Simulation					
Simulation					
S/W Development					
S/W Development					
S/W Development					
Other (7)					
Other					

Cobol Projects

Effort, SLOC, and Sch

Number of records included in search: 13

Application	Tot Eff	Norm Eff Sz	Sch Mon	Prog Lang
MIS (1)	1833.00	419619	30	Assembly 1% COBOL 65% Other 34%
MIS	3960.00	419619	50	Assembly 1% COBOL 61% Other 38%
MIS (2)	736.00	97087	45	COBOL 82% Other 18%
MIS	2574.00	461426	60	COBOL 34% Other 66%
MIS	1115.00	231018	26	Basic 5% C 10% COBOL 46% Other 39%
MIS	1525.00	363371	58	C 2% COBOL 47% Other 51%
MIS	1167.00	200000	22	COBOL 50% Other 50%
MIS (3)	202.00	6681	36	COBOL 100%
MIS (4)	225.00	7457	36	COBOL 100%
MIS (5)	652.00	21588	36	COBOL 100%
MIS (6)	439.00	14536	36	COBOL 100%
MIS (7)	358.00	11840	36	COBOL 100%
MIS (8)	299.00	9899	36	COBOL 100%

Cobol Projects

Function Points

Number of records included in search: 13

Application	Ext Inpts	Ext Outpts	Ext Inq	Int Files	Ext Int
MIS (1)					
MIS					
MIS (2)					
MIS					
MIS					
MIS					
MIS					
MIS (3)					
MIS (4)					
MIS (5)					
MIS (6)					
MIS (7)					
MIS (8)					

Missile

Function Points

Number of records included in search: 5

Application	Ext Inpts	Ext Outpts	Ext Inq	Int Files	Ext Int
Command/Control	18	17	0	161	1
Command/Control	18	17	0	108	1
OS/Executive	32	17	17	22	2
Command/Control					
Command/Control	15	11	4	3	6

Missile

Effort, SLOC, and Sch

Number of records included in search: 5

Application	Tot Eff	Norm Eff Sz	Sch Mon	Prog Lang
Command/Control	276.00	8885		Assembly 100%
Command/Control	96.00	9025		Assembly 100%
OS/Executive	77.00	1002		Assembly 100%
Command/Control	1460.00	18933		JOVIAL 100%
Command/Control	506.00	13658		Assembly 100%

Bibliography

Boehm, B.W. Software Engineering Economics. Englewood Cliffs NJ: Prentice-Hall, Inc., 1981.

-----. "Software Engineering Economics," IEEE Transactions on Software Engineering, 1:239- 256 (1984).

Coggins, G.A. & R.C. Russell. Software Cost Estimating Models: A Comparative Study of What the Models Estimate. MS thesis, AFIT/GCA/LAS/93S-4. School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, 1993 (AD-A275989).

Conte, S.D., H.E. Dunsmore and V.Y. Shen. Software Engineering Metrics and Models. Menlo Park CA: The Benjamin/Cummings Publishing Company, Inc., 1986.

Christensen, D.S. & D.V. Ferens. "Software Cost Model Calibration - An Air Force Case Study," School of Logistics and Acquisition Management, Air Force Institute of Technology, Wright-Patterson AFB OH.

DeMarco, T. Controlling Software Projects. Englewood Cliffs NJ: Prentice-Hall, 1982.

Fairley, R.E. Software Engineering Concepts. New York: McGraw-Hill Book Company, 1985.

Ferens, D.V. Class handout, COST 677, Quantitative Management of Software, School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, Fall Quarter 1995.

Ferens, D.V. & R.B. Gurner. "An Evaluation of Three Function Point Models for Estimation of Software Effort," School of Logistics and Acquisition Management, Air Force Institute of Technology, Wright-Patterson AFB OH, 1994.

Fisher, G.H. "A Discussion of Uncertainty in Cost Analysis," Memorandum RM-3071-PR, Contract AF 49(638)-700. Santa Monica CA: The RAND Corporation, 1962.

Galonsky, J.C. Calibration of the PRICE S Software Cost Model. MS Thesis, AFIT/GCA/LAS/95S-1. School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1995 (AD-A301377).

Hayes, John. Cost Analyst, Wright-Patterson AFB OH. Personal Interview. 2 February 1996.

Illinois Institute of Technology (IIT) Research Institute, Test Case Study: Estimating the Cost of Ada Software Development, Lanham MD, IIT: April 1989.

Jones, C. "Software Metrics: Good, Bad, and Missing," Computer, 27: 98-100, September, 1994.

Kressin, R.K. Calibration of SLIM to the Air Force Space and Missile Systems Center Software Database. MS Thesis, AFIT/GCA/LAS/95S-6. School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1995 (AD-A301603).

Londeix, B. Cost Estimating for Software Development. New York: Addison-Wesley Publishing Company, 1987.

Mendenhall, W., D. Wackerly, and R. Scheaffer. Mathematical Statistics With Applications (Fourth Edition). Belmont CA: Duxbury Press, 1990.

Novak-Ley, G. and S. Stukes. SMC SWDB User's Manual: Version 2.1. Oxnard CA: Management Consulting & Research, Inc., 1995.

Ourada, G.L. and D.V. Ferens. "Software Cost Estimating Models: A Calibration, Validation, and Comparison,." Cost Estimating and Analysis: Balancing Technology and Declining Budgets 83-102, 1991.

Rathmann, K.D. Calibration and Evaluation of SEER-SEM for the Air Force Space and Missile Systems Center. MS Thesis, AFIT/GCA/LAS/95S-9. School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1995 (AD-A300703).

Software Productivity Research. Checkpoint for Windows User's Guide. Burlington MA, September 1993.

Stukes, Sherry. Senior Associate for Management Consulting and Research, Oxnard CA. Personal Interview. 2 April 1996.

Thibodeau, R. "An Evaluation of Software Cost Estimating Models." Huntsville AL: General Research Corporation, 1981.

Tinkler, Shirley. Cost Analyst, Space and Missile Systems Center, Los Angeles CA. Personal Interview. 2 April 1996.

Vegas, C.D. Calibration of Software Architecture Sizing and Estimation Tool. MS Thesis, AFIT/GCA/LAS/95S-11. School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1995 (AD-A301376).

Weber, B.G. A Calibration of the REVIC Software Cost Estimating Model. MS Thesis, AFIT/GCA/LAS/95S-13. School of Logistics and Acquisition Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1995 (AD-A300694).

Wellman, F. Software Costing: An Objective Approach to Estimating and Controlling the Cost of Computer Software. New York: Prentice-Hall, Inc., 1992.

Zimmerman, John. Senior Consultant for Software Productivity Research, San Francisco CA. Personal Interview. 20 May 1996.

Vita

Capt Karen R. Mertes [redacted]

She graduated from Shrewsbury High School in 1984 and entered undergraduate studies at Boston University in Boston, Massachusetts. She graduated with a Bachelor of Arts degree in Mathematics in May 1988. She received her commission as a distinguished graduate through Air Force ROTC on 13 May 1988.

Upon completion of Intelligence Applications Officer School in December 1989, Capt Mertes was assigned as Chief, Interrogation and Wartime Training at Fort Belvoir, Virginia. In August 1991, she was assigned as a Human Resources Intelligence Collector at Fort Belvoir, Virginia. In September 1992, she transferred to Andrews AFB, Maryland as Chief of Mission Support. Capt Mertes received a Master of Science in Business Administration degree from Strayer College in 1995. She entered the AFIT Graduate Cost Analysis program in May 1995. On September 24, 1996 she graduated with a Master of Science in Cost Analysis. Her follow on assignment was Kirtland AFB, New Mexico.

[redacted]

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1996	Master's Thesis

4. TITLE AND SUBTITLE
CALIBRATION OF THE CHECKPOINT MODEL TO THE SPACE AND MISSILE SYSTEMS CENTER (SMC) SOFTWARE DATABASE (SWDB)

5. FUNDING NUMBERS

6. AUTHOR(S)
Karen R. Mertes, Captain, USAF

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

Air Force Institute of Technology
2750 P Street
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GCA/LAS/96S-11

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

Management Consulting & Research, Inc.
4165 E. Thousand Oaks Boulevard, Suite 235
Thousand Oaks, CA 91362

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 Words*)

This thesis analyzed the effect of calibration on the performance of the CHECKPOINT Version 2.3.1 software cost estimating model. Data used for input into the model were drawn from the FY 95 USAF SMC Software Database (SWDB) Version 2.1. A comparison was made between the model's accuracy before and after calibration. This was done using records which were not used in calibration, referred to as validation points. A comparison of calibration points, both before and after, was done in order to assess whether calibration results in more consistency within the data set used. Six measures such as magnitude of relative error (MRE), mean magnitude of relative error (MMRE), root mean square error (RMS), relative root mean square error (RRMS), the prediction at level *k/n,* and the Wilcoxon Signed-Rank Test were used to describe accuracy. The results of this effort showed that calibration of the CHECKPOINT model can improve cost estimation accuracy for development effort by as much as 96.71%.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Software, Computers, Computer Programs, Software Engineering, Calibration, Models, Cost Models, Cost Estimates	120
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED

AFIT Control Number AFIT/GCA/LAS/96S-11

AFIT RESEARCH ASSESSMENT

The purpose of this questionnaire is to determine the potential for current and future applications of AFIT thesis research. **Please return completed questionnaire** to: AIR FORCE INSTITUTE OF TECHNOLOGY/LAC, 2950 P STREET, WRIGHT-PATTERSON AFB OH 45433-7765. Your response is **important.** Thank you.

1. Did this research contribute to a current research project? a. Yes b. No

2. Do you believe this research topic is significant enough that it would have been researched (or contracted) by your organization or another agency if AFIT had not researched it?
a. Yes b. No

3. **Please estimate** what this research would have cost in terms of manpower and dollars if it had been accomplished under contract or if it had been done in-house.

Man Years_____ $_____

4. Whether or not you were able to establish an equivalent value for this research (in Question 3), what is your estimate of its significance?

a. Highly Significant b. Significant c. Slightly Significant d. Of No Significance

5. Comments (Please feel free to use a separate sheet for more detailed answers and include it with this form):

_____ _____
Name and Grade Organization

_____ _____
Position or Title Address