2022-05-02

Improving Software Cost Estimating Techniques in Defense Programs

Biancalana, Nickolas A.; Ritschel, Jonathan D.; Drylie, Scott T.; Fass, David; White, Edward D.

Monterey, California. Naval Postgraduate School

SYM-AM-22-077

Excerpt from the Proceedings

of the

Nineteenth Annual Acquisition Research Symposium

Acquisition Research:
Creating Synergy for Informed Change

May 11–12, 2022

Published: May 2, 2022

The research presented in this report was supported by the Acquisition Research Program at the Naval Postgraduate School.

Improving Software Cost Estimating Techniques in Defense Programs

First Lieutenant Nickolas A. Biancalana—is a Cost Analyst at the Air Force Cost Analysis Agency at Joint Base Andrews, MD. He received his BS in Business Management from the United States Air Force Academy and his MS in Cost Analysis from the Air Force Institute of Technology (AFIT). His research interests include strategic decision making, optimization, and data analytics. [nickolas.biancalana.1@us.af.mil]

Dr. Jonathan D. Ritschel—is an Assistant Professor of Cost Analysis in the Department of Systems Engineering and Management at AFIT. He received his BBA in Accountancy from the University of Notre Dame, his MS in Cost Analysis from AFIT, and his PhD in Economics from George Mason University. Ritschel's research interests include public choice, the effects of acquisition reforms on cost growth in DoD weapon systems, and economic institutional analysis. [jonathan.ritschel@afit.edu]

Lieutenant Colonel Scott T. Drylie, PhD—is an Assistant Professor of Cost Analysis in the Department of Systems Engineering and Management at AFIT. He received his BS in Economics from Montana State University, his MS in Cost Analysis from AFIT, and his PhD in Economics from George Mason University. His research interests include institutional economics, the economics of education, software costs, and aircraft sustainment costs. [scott.drylie@afit.edu]

Dr. R. David Fass—is an Assistant Professor of Systems Integration and Cost Analysis in the Department of Systems Engineering and Management at AFIT. He received his BA in Economics and English from the University of New Mexico, his MBA from the University of New Mexico, and his PhD in Philosophy, Business Administration & Management from New Mexico State University. Fass's research interests include cost analysis, decision analysis, and corporate and business strategies. [robert.fass@afit.edu]

Dr. Edward D. White—is a Professor of Statistics in the Department of Mathematics and Statistics at AFIT. He received his BS in Mathematics from the University of Tampa, MAS from The Ohio State University, and PhD in Statistics from Texas A&M University. His primary research interests include statistical modeling, simulation, and data analytics. [edward.white@afit.edu]

Abstract

As software becomes more ubiquitous in defense programs, there is a need to improve the accuracy and reliability of methods for estimating software size and cost. Historically, practitioners have used defined distributions in their estimating software to simulate likely outcomes. This research identifies new distributions of likely software costs and effective sizes through an analysis of Cost and Software Data Reports (CSDRs) as well as demonstrating the most appropriate distribution given certain program characteristics known at the genesis of the project. By utilizing various descriptive statistics and statistical tests, this research shows there are distributions that are more closely tailored to the actual qualities of a software program. In some instances, a broad and general distribution is sufficient; however, there are specific commodities, contractors, and system types that are distinctly different and require additional analysis. Overall, this research intends to equip practitioners with an arsenal of distributions and statistical information that will lead them to apply the best model to predict software size and cost, all with the goal of improving overall accuracy.

Introduction

Software has become a core functional element in many defense projects and therefore plays a vital role in the definition of mission critical capabilities (McQuade et al., 2019). For that reason, it is prudent to utilize a consistent and accurate method to properly quantify the expected costs associated with incorporating software into defense projects. Given the implications of improperly estimating software systems and, in turn, the entire project, (e.g., cost overruns, inadequate funding, etc.) accurate estimates are paramount. The increasingly robust centralized cost databases (e.g., Cost Assessment and Data Enterprise [CADE]) provide the opportunity for cost analysts to access a wide array of historical software data points. This data can be used to create new distributions that form accurate regions of reasonable estimates, ultimately helping the cost analyst perform a more precise estimate.

By analyzing historical defense projects from various branches in the military, this research seeks to identify patterns between different characteristics of projects and how they relate to the final cost of software packages. Once these relationships are uncovered, that information can shed light on how to properly size software to new projects. The software data from these historical projects holds the key to refining the estimating process. Additionally, it will provide practitioners with an arsenal of distributions that can be used as inputs into simulation software to create likely outcomes for the cost and size of software programs. This will increase confidence in estimates as well as provide clarity as to how software costs relate to the mission in which they are designed to serve. The implications include not only formulating more accurate estimates but also knowing what a realistic cost would be prior to accepting contractor proposals. This puts the DoD in an advantageous and leveraged position during negotiations while also mitigating potential risk of cost overruns.

Background

Considering the array of possible applications of software and the platforms in which they serve, it would be unreasonable to consider the software programs and their costs homogenous. The ability to obtain a more specific distribution of the likely costs associated with software in a project given various characteristics of the project itself is vital. The efforts of Sheppard and Schofield (1997) show that analogy methods predict software exceptionally well compared to regression-based analysis. By creating more comprehensive analogies and specific distributions, estimators will have a more refined tool to formulate accurate and precise estimates. Additionally, it will educate decision makers as to what is and is not a reasonable contractor proposal. Previous research regarding software systems in DoD programs has been conducted; however, it involved software size as it related to effort levels rather than the cost itself (Madachy et al., 2011; Sheppard & Schofield, 1997).

This research aims to explore a new and unique angle to software costs in defense projects. This research is unique in that it is looking at the individual costs of the software itself within the defense programs and using those to create comprehensive analogies to aid in future cost estimates. Up to this point, this approach has not been pursued and data had not been readily available. The Air Force Life Cycle Management Center (AFLCMC) collected and provided consolidated datasets containing not only the software characteristics of dozens of different projects, but also each project's respective cost information. This was accomplished by taking the Software Resource Data Reports (SRDR) for each project from CADE and matching those software characteristics with the cost information found on the project's Cost Data Summary Report (i.e., Form 1921). This data included information regarding the commodity, branch of service, nonrecurring costs, total lines of code, Effective Source Lines of Code (ESLOC), team structure with regards to experience level, number of hours in each phase of development and a multitude of other measurements. The dataset is among the first of its kind in that it combines a project's software data (lines of code, primary language, etc.) with its cost data, allowing for a comprehensive analysis of the relationships between cost and software as it relates to difference project characteristics.

Problem Statement/Research Questions

One of the problems this research addresses is accounting for diversity in software. There is a wide variety of defense projects, spanning an enormous range of software specifications, software requirements, and ultimately software costs. From a cost analyst's point of view, this makes estimating a software system's cost a particularly perilous task. If there were a way to narrow the range of possible values given certain characteristics of a project, the analyst would be able to provide a more accurate and confident estimate of a project's software costs. To address this problem, this paper examined the question, "How do the size and cost of software packages relate to the project in which they operate, and how do they change as the characteristics of the project are changed?

Literature Review

The motivation behind this literature was to validate or contradict the selected elements of this research. At any level of this literature review, if the elements were invalidated or found to be of little use, the purpose of the research would be of little use. Starting at the top level, software in and of itself is being increasingly relied upon in the DoD (GAO, 2021). Considering software is now at the forefront of DoD acquisition programs either in a direct or supporting role, the methods and techniques used to estimate the costs must be fortified.

Next, utilizing the results of this research implies the use of analogy and parametric estimating methods. The analogy method entails finding an analogous program and scaling its parameters to model the new program based on its known characteristics. This method has several advantages and disadvantages (Garrett, 2008; Kueng, 2008). The main disadvantages stem from the analogy itself and its appropriateness. If the analogy cannot be defended and should not be used, the estimate created has lost its value. Despite the disadvantages, this method has shown to be superior to regression-based estimating within the realm of software (Sheppard & Schofield, 1997). The parametric method involves using parametric models that have been derived from cost driving factors that are found by developing statistical relationships between historical costs and program, physical, and performance characteristics (Garrett, 2008). This method also has its advantages and disadvantages (Pfleeger et al., 2005) and has been refined through decades of research. AFLCMC uses a form of parametric modelling in their software cost estimations as well. They input known distributions for various project parameters into their estimating software and perform simulations. These simulations result in distributions for overall costs that are used in the decision-making process.

Next, regarding the independent variables of this research, previous works have segmented datasets into groups that resemble the groups used in this research. Jones et al. (2014), although investigating a common rule of thumb in O&S cost estimating, segments their dataset into groups labeled Space, Fixed-Wing Aircraft, Rotary-Wing Aircraft, Missiles, Electronics, Ships, Surface Vehicles, and Automated Information Systems (AIS). They further segment Fixed-Wing Aircraft into Fighter, Cargo/Tanker, and Unmanned Aerial Vehicles (UAV). These groups very closely resemble the commodity and system type groups used in this research. Their results showed the need to segment projects by these categories and found differences between them (Jones et al., 2014). Additionally, as part of their statistical analysis, Madachy and Clark (2015) segmented their data by "operating environment." Members of this group included Aerial Vehicle (including fixed-wing, rotary-wing, and unmanned aircraft), Space Vehicle, and Ordnance Vehicle (including missiles). Simultaneously, given the missions of each project, each article is also segmenting their datasets by Service, although not explicitly. These sources show an intuition to separate projects and create homogenous groups such as commodity and system type and explain it is unwise and imprudent to treat all projects the same.

Lastly, the use of these specific dependent variables must be validated. This research utilized Effective Source Lines of Code (ESLOC) and nonrecurring costs as a rate of ESLOC. The Air Force Cost Analysis Agency (AFCAA) and Naval Center for Cost Analysis (NCCA) describe effective size as a major factor of software cost and schedule estimating (AFCAA, 2008). AFCAA goes on to explain the role of Effective Source Lines of Code (ESLOC) and how it relates size to work. They explain,

> Resource estimates based on physical source lines of code for modified software and systems containing reusable components, cannot account for the additional resource demands attributable to reverse-engineering, test, and software integration. The common method used to account for the added resource demands is to use the *effective* software size. (AFCAA, 2008)

Additionally, Clark and Madachy further this statement in the Software Cost Estimation Metrics Manual for Defense Systems (2015) and state equivalent size is "a key element in using software size for effort estimation" (Clark & Madachy, 2015). They go on to assert that equivalent size quantifies how much effort is required to reuse old code alongside new code. ESLOC is a pivotal measurement that encapsulates both size and complexity.

With the introduction of cost, it is important to distinguish between recurring and nonrecurring costs. The Defense Acquisition University defines nonrecurring costs as "costs that will occur once or occasionally for a particular cost objective, NRCs include preliminary design effort, design engineering, and all partially completed reporting elements manufactured for tests" (DAU Glossary, n.d.). Additionally, they describe a recurring cost as "costs for items and services that reoccur, especially at regular intervals. Recurring costs are incurred each time a unit equipment is produced, such as direct labor and direct materials" (DAU Glossary, n.d.). Since the costs that this research is focused on is the preliminary design and engineering of software packages, nonrecurring costs will be assessed in the form of the rate nonrecurring cost per ESLOC.

Research Gap

Previous research has looked at past relationships between software size, effort, productivity, and complexity, but normalized historical costs have not been included in the analysis. This current research is not only aimed at utilizing previous costs to establish relationships and distributions to predict future costs but also investigating program characteristics and how they influence key cost drivers such as ESLOC.

The data from the CADE database directly links the software characteristics from a program's Software Resources Data Report to its cost data from its Form 1921. Previous research has investigated software through various lenses; however, this dataset finally allows for the direct analysis of past costs and not relying on some form of a proxy to estimate costs. This offers the opportunity to poignantly investigate distributions regarding both the cost per ESLOC parameter and the ESLOC parameter itself. As stated earlier, ESLOC is a key metric in software models that encapsulates the size and effort of a project. Additionally, creating a rate of cost per ESLOC standardizes each project in the dataset to avoid distortions from exceptionally large and/or expensive projects.

Data

The data used in this research is a combination of datasets from AFLCMC and CADE. AFLCMC provided the consolidated data containing program characteristics, software components and capabilities, nonrecurring costs, and many other quantities and dates pertaining to the development and purchase of the software packages. This dataset contained 44 different programs across the DoD with detailed information down to the WBS element. This dataset is a consolidation of the WBS element's software characteristics and properties found on the project's Software Resource Data Reports (SRDR) and the element's cost information found on the project's DD Form 1921. This data was collected by AFLCMC from CADE and consolidated for this research. The AFLCMC data was verified by taking a 10% (50 WBS elements) sample and comparing the information to the sourced data from CADE. Once the sample was taken, each WBS element from the AFLCMC data was found in the CADE dataset and compared for accuracy. Of the 50 WBS elements used, all matched the CADE dataset giving confidence there are few or no mismatches in the AFLCMC data.

Inclusion Criteria

Now that there is a consolidated and verified dataset, lines containing outliers or missing data must be excluded from the analysis. Due to the highly skewed nature of this dataset, a more traditional outlier test such as three times the Interquartile Range (IQR) beyond the 25th and 75th percentiles was not practical. The skew present in the data caused the IQR to be very small which would place the outlier bounds closer to the median. If this approach were taken, nearly 10% of this dataset would be excluded. For this research, initial outliers were identified using a quantile range exclusion method (Klimberg & McCollugh, 2016). This method calculated the range from the fifth to 95th percentiles, multiplied this range by three and excluded any data points beyond that distance from the fifth and 95th percentiles. For example, with regards to Nonrecurring Cost/ESLOC, the calculation is as follows:

$$Range = 95th\ Percentile - 5th\ Percentile \tag{1.1}$$
$$Range = \$3.767K - \$0.009K = \$3.758K$$

$$Outlier\ Lower\ Bound = 5th\ Percentile - (Range * 3) \tag{1.2}$$
$$Outlier\ Lower\ Bound = \$0.009K - \$11.274K = -\$11.265$$

$$Outlier\ Upper\ Bound = 95th\ Percentile + (Range * 3) \tag{1.3}$$
$$Outlier\ Upper\ Bound = \$3.767K + \$11.274K = \$15.041K$$

With these bounds now established, any observations beyond them were excluded from the dataset. Since the lower bound was negative and neither Cost per ESLOC nor ESLOC can be negative, the values were truncated at zero. This technique was performed for both the Cost/ESLOC analysis as well as the ESLOC analysis. For the rate analysis, three observations were removed and for the ESLOC analysis, two were removed.

Table 1. Distribution Analysis Sample Sizes with Exclusions

	Cost/ESLOC Analysis	**ESLOC Analysis**
Total Initial Data Points	460	460
Missing Values	106	66
Outliers	3	2
Data Points Remaining (all analyses except contract type)	**351**	**392**
No viable contract information	37	38
Data Points Remaining (contract type analysis only)	**314**	**354**

Methodology

The overall approach for this research was a process dubbed "incremental analysis." The purpose of incremental analysis is to observe how the dependent variables (Cost/ESLOC and ESLOC) change as other variables are changed. Traditionally, a regression model would show the individual effects of each independent variable on the dependent variable. However, there were many interactions between independent variables within this dataset that would decrease the overall utility of the model. If a regression model were pursued, the outcome would contain many interactions variables pertaining to specific combinations of contractor and commodity, commodity and service, contract type and commodity, and so on. The resulting regression model would indicate effects on the dependent variables; however, they would only apply to those specific combinations and would lack utility.

The alternative is to do a series of bivariate analyses with various combinations of independent variables to observe how the dependent variable changes. Additionally, these analyses would show which individual combinations are different from one another thus identifying variables that have more impact on the dependent variables than others. These unique differences also illuminate which combinations of independent variables require a distribution of their own outside of the univariate distributions found for each individual variable.

The incremental analysis was performed twice for each pair of independent variables. One analysis for a given pair of characteristics holds one independent variable constant while varying the other and the second analysis switches the variables. Within each analysis, the median cost ($K/ESLOC) and effective size (KESLOC) is reported. This is due to the skewed nature of the data and as a result, a mean would not be a good representation for the data. Given there are five program characteristics in this research, there are 25 total combinations. This method is repeated for each combination of independent variables except for the combination of commodity and system type since system type is a subgroup of commodity. Additionally, a characteristic will not be compared against itself. After these removals, there were 18 total combinations explored in this research. All 18 combinations will be explored for both dependent variables, Cost per ESLOC and ESLOC. For the purposes of this paper, only those pertaining to a project's commodity are discussed.

Each analysis contains a Kruskal-Wallis p-value which compares the values within the constant variable as it's changed by the other variable. This p-value indicates whether differences are detected between the values and the subsequent Steel-Dwass test highlights which specific pairs of values are different from one another. An alpha level of 0.05 was used for both tests. The Steel-Dwass outcomes for each pair of analyses are then compared and any overlaps in results indicate a specific combination of variables that warrants its own distribution. This is because two Steel-Dwass tests have shown that each variable that is part of the specific combination was different than at least one of the other categories within its subset.

These specific combinations were then fit with multiple probability density functions (PDFs), and each was evaluated on how well it fit the distribution. Due to the practitioner's familiarity, lognormal distributions were always provided, regardless of whether it was the best fit or not. The Anderson-Darling test result is provided so the practitioner is aware if a lognormal distribution is not an appropriate method to model this data and should use the better fitting distribution.

Analysis and Results

This section contains incremental analyses showing how Cost/ESLOC and ESLOC change when one variable varies and another is held constant, all in search of more specific combinations of variables that warrant a separate distribution.

Cost Per ESLOC Analyses

The following analysis identifies how Cost/ESLOC changes as various independent variables are changed. Each iteration of this analysis will take two independent variables, hold one constant, and assess how the median values of Cost/ESLOC change as the Other independent variable is changed. The variables are then switched regarding which is held constant to identify any unique pairs of variables that warrant a deeper analysis.

Cost/ESLOC—Commodity and Contractor

Table 2 illustrates a two-way dissection of the Cost/ESLOC rate. It segments the data first by contractor, then by commodity. It also shows the differences between commodities within the same contractor. The numbers within the table represent the median Cost/ESLOC in thousands for each commodity within each contractor. The bottom three rows of the table show the total number of observations, median value, Kruskal-Wallis p-value for the test performed on the commodities within a certain contractor. Steel-Dwass pairs are annotated by shared letters in the cells.

Table 1. Contractor by Commodity Analysis—Cost ($K)/ESLOC

Commodity/Contractor	Contractor 1	Contractor 2	Contractor 3	Contractor 4	Contractor 5
Aircraft	0.183	0.185	0.174	0.088[b]	0.043[b]
EAS	0.278	0.177[a]	0.056	0.134[a]	0.071[a]
Missile		0.708			0.511[a,b,c]
Rotary Wing	0.202	0.467	0.218	0.327	
Space	0.798	1.331[a]	0.131	1.051[a,b]	0.078
UAV			0.173		0.141[c]
N	96	46	89	52	68
Median	0.205	0.377	0.119	0.28	0.122
KW	0.303	0.004	0.05	0.001	0.008

Note: Commodities that share a letter within the same contractor are members of a Steel-Dwass Pair

The correct way to interpret Table 2 is as follows. The Kruskal-Wallis p-value (0.303) is not smaller than 0.05, meaning there is not sufficient evidence to say any of the commodities within Contractor 1 are different from one another.

The same results are found when looking at Contractor 3 in that none of the commodities are distinctly different. However, Contractors 2, 4, and 5 all have significantly low Kruskal-Wallis p-values, indicating that there are differences between commodities within the single contractors. Additionally, one can determine which contractor produces more expensive commodities by comparing the median values of each member of a Steel-Dwass pair.

The results of this analysis show that even within a singular contractor, differences can be found between commodities. Additionally, these results show that while holding contractor constant, the Cost/ESLOC changes as commodity changes. Prior to this analysis, the contractor or commodity was analyzed at large but this shows that even within a particular contractor, further analysis may still be required to find the most appropriate distribution.

This analysis only represents one side of this investigation. Although this analysis showed differences within a specific contractor, if a difference cannot be found between contractors within the same commodity, then one would be better off to use the overall commodity distribution. However, if a difference is found between commodities within the same contractor and that same difference is found between contractors within the same commodity, an even more specific distribution would be required.

Table 3. Commodity by Contractor Analysis—Cost ($K)/ESLOC

Contractor/Commodity	Aircraft	EAS	Missile	RW	Space	UAV
Contractor 1	0.183	0.278[a,b]		0.202	0.798	
Contractor 2	0.185	0.177	0.708	0.467	1.331[b]	
Contractor 3	0.174	0.056[a]		0.218	0.131[a,b]	0.173
Contractor 4	0.088	0.134		0.327	1.051[a]	
Contractor 5	0.043	0.071[b]	0.511		0.078	0.141
N	75	105	21	61	38	51
Median	0.149	0.127	0.577	0.207	0.57	0.143
KW	0.308	0.001	0.622	0.334	0.005	0.434

Note: Contractors that share a letter within the same commodity are members of a Steel-Dwass Pa

Table 3 shows the same median values, but this analysis switches the rows and columns compared to the previous analysis. This analysis now shows the differences between contractors within the same commodity. Again, the Kruskal-Wallis p-values show whether significant differences were detected within a given commodity. This test found significance in the EAS and Space commodities. Contractors that share a letter within the same commodity are members of a Steel-Dwass pair and their relationship to one another can be found my comparing median values.

Like the previous one, this analysis identifies differences between contractors within the same commodity as well as which produces a more or less expensive software component than another. This shows that segmenting the data by only commodity may still not be sufficient given the differences that were found.

If there are overlapping differences in the preceding analyses, additional distributions are required. For example, looking at the Steel-Dwass pairs from the Contractor by Commodity analysis, Contractor 2 has a Steel-Dwass pair of Space and EAS. This means that within Contractor 2, Space and EAS are distinctly different from one another. Knowing this information and looking at the Commodity by Contractor analysis, if Contractor 2 is a member of a Steel-Dwass pair for either Space or EAS, that would require a new distribution since both components have been shown to be distinctly different. This instance occurs with Contractor 2 Space programs. Simply put, Contractor 2 Space programs have shown to be different than other commodities that Contractor 2 works on and different than space programs that other contractors do. This warrants an additional distribution due to the dual differences found. This distribution is shown below.

Table 4. Contractor 2/Space Distributions—Cost ($K)/ESLOC

	Median	Distribution	AD
Contractor 2 - Space	1.331	Exponential	0.968
		Lognormal	0.879

Table 4 shows the distribution information for all data points representing Space programs accomplished by Contractor 2. An Exponential distribution is the best fit based on AIC and the p-value indicates that it is appropriate to use an exponential distribution to model this information. Additionally, a lognormal distribution would also be appropriate given the parameters above. Since this dataset is small it is difficult to draw firm conclusions regarding the overall shape of these data points but given the data at hand, these distributions would be appropriate. This phenomenon occurs twice more, the first being Contractor 4 (mostly subcontractors or contractors with few data points) and Space.

Table 5. Contractor 4/Space Distributions—Cost ($K)/ESLOC

	Median	Distribution	AD
Contractor 4 - Space	1.051	Exponential	0.758
		Lognormal	0.987

Like the previous distribution, both an exponential and lognormal distribution would be an appropriate fit. Again, due to a small sample size, it is difficult to draw firm conclusions; however, based upon the data at hand, these distributions would be appropriate. Lastly, this occurs again with Contractor 5 and EAS.

Table 6. Contractor 5/EAS Distributions—Cost ($K)/ESLOC

	Median	Distribution	AD
Contractor 5 - EAS	0.071	Exponential	0.179
		Lognormal	0.293

These three distributions represent the most specific and detailed level that retains relevancy. Other than these, the lowest level required would be either the contractor level or commodity level. However, since the components of these three distributions have shown to be different than their counterparts', more detailed distributions are required. The remaining analyses were performed in the same manner as outlined above but their results are presented in an abridged format.

Cost/ESLOC—Commodity and Service

The next analysis held commodity constant while varying service. After performing the statistical tests, it was found that only the EAS commodity had significant differences between services. The Army was found to be significantly cheaper than both the Air Force and Navy.

Next, commodity was varied within each service. Both the Air Force and Army had significant differences detected. Within the Air Force, Space was found to be more expensive than both Aircraft and EAS. Within the Army, EAS was found to be cheaper than Missile and Rotary Wing. These results coupled with the results of the previous analysis show there are two instances where a more specific distribution is required.

Table 7. Air Force/EAS Distribution—Cost ($K)/ESLOC

	Median	Distribution	AD
Air Force - EAS	0.192	Lognormal	0.764

The table above shows the best distribution for the specific combination of Air Force and EAS. Based on AIC, a lognormal PDF was the best fit for this distribution and based on its Anderson-Darling p-value, it is also an appropriate method to model data with these characteristics.

Table 8. Army/EAS Distribution—Cost ($K)/ESLOC

	Median	Distribution	AD
Army - EAS	0.057	Lognormal	0.523

The other overlap occurs again with EAS but this time with the Army. Again, lognormal was the best fit based on AIC and is appropriate based on Anderson-Darling p-value.

Cost/ESLOC—Commodity and Contract Type

The next analysis investigates how Cost/ESLOC changes when contract type is held constant, and commodity is changed. None of the specific contract types contained significant differences between commodities except for Mixed Contracts (MC). Space was found to be more expensive than Rotary Wing, EAS, Aircraft, and UAV. Given there is no clean definition as to what exactly comprises a Mixed Contract, it is difficult to draw any firm conclusions from these results.

Next, commodity and contract type were switched to investigate how Cost/ESLOC changed when commodity is held constant and contract type is varied. Only Aircraft and EAS had detectable differences. Within Aircraft, CPFF contracts were cheaper than Mixed Contracts. Within EAS, CPFF was found to be cheaper than CPAF, CPIF, CW, and MC.

There are two specific combinations of commodity and contract type that appears in both sets of Steel-Dwass pairs and warrants a more specific distribution. The first distribution is shown below.

Table 9. MC/EAS Distribution—Cost ($K)/ESLOC

	Median	Distribution	AD
MC - EAS	0.079	Lognormal	0.152

Table 9 shows the best fitting PDF for EAS commodity WBS elements performed on a Mixed Contract. Based on AIC, lognormal was the best fit and based on the Anderson-Darling p-value, it is also an appropriate fit due to the value being larger than the alpha level of 0.05. It is difficult to put the utility of this distribution in perspective since there is no clear definition of a Mixed Contract in terms of composition of fixed versus cost-plus elements.

Table 10. MC/Aircraft Distributions—Cost ($K)/ESLOC

	Median	Distribution	AD
MC - Aircraft	0.259	Exponential	0.372
		Lognormal	0.026

Table 10 shows the best fitting distribution for Aircraft commodities using a mixed contract type. Exponential was the best fit based on AIC and the Anderson-Darling p-value shows it is appropriate to use. Lognormal is provided but the p-value shows it is not an appropriate PDF to use to model this data.

ESLOC Analyses

The following analyses are identical in nature to the Cost/ESLOC analyses shown previously except now the dependent variable is ESLOC in thousands. These results will show how ESLOC changes when one independent variable is held constant and the Other is changing, illuminating the impacts of these independent variables.

ESLOC—Commodity and Contractor

Like the previous contractor by commodity analysis performed on Cost/ESLOC, this analysis showed how ESLOC changes when the commodity is changed, all while contractor is held constant. Only Contractors 1 and 2 had detectable differences between commodities. Within Contractor 1, Aircraft projects showed to lead to a significantly larger effective size than Rotary Wing. Within Contractor 2, Missile was found to be smaller than EAS and Space.

Next, commodity was held constant while contractor was varied. Only within the Missile commodity were differences found. Contractor 2 was found to create much smaller software programs than Contractor 5.

One overlap occurred with Contractor 2's Missile projects. The combination of Contractor 2's Missile projects being different than other commodities in which they have performed work and different than other contractors' Missile projects warrant a separate distribution to model this specific relationship.

Table 11. Contractor 2/Missile Distributions—ESLOC (K)

	Median	Distribution	AD
Contractor 2 - Missile	4.948	Exponential	0.694
		Lognormal	0.004

The exponential distribution was the best fit for this data based on AIC and is appropriate due to the Anderson-Darling p-value. The lognormal distribution on the other hand is not an appropriate tool to model this data since its p-value is below the alpha level of 0.05.

ESLOC—Commodity and Service

The next two analyses investigate the impact on ESLOC (K) when commodity and service are changed. The first changes service while holding commodity constant. Only within the EAS, Missile, and UAV commodities were differences detected. Within the EAS commodity, yielded significantly larger effective sizes than both the Air Force and Navy. Within the Missile commodity, The Army was significantly smaller than the Air Force. Lastly, within the UAV commodity, the Navy was significantly larger than the Air Force.

Next, the previous independent variables are switched, and service is held constant while commodity is varied. Within the Army, Rotary Wing was found to be larger than Missile, and EAS was found to be larger than both Missile and Rotary Wing. Within the Navy, EAS was found to be smaller than Aircraft.

There are three instances where the Steel-Dwass pairs overlap and require a more specific distribution.

Table 12. Army/Missile Distributions—ESLOC (K)

	Median	Distribution	AD
Army - Missile	4.948	Exponential	0.649
		Lognormal	0.006

The first overlap occurs with Missile WBS elements performed by the Army. The table above shows exponential as the best fitting PDF based on AIC. Due to the practitioner's familiarity with lognormal distributions, it is also provided. Based on Anderson-Darling p-values, the exponential distribution is an appropriate method to model this data since the p-value is larger than the alpha level of 0.05. However, a lognormal distribution is not an appropriate method because the p-value is less than the alpha level.

Table 13. Army/EAS Distribution—ESLOC (K)

	Median	Distribution	AD
Army - EAS	214.071	Lognormal	0.195

The next overlap occurs with the Army and the EAS commodity. Lognormal was the best fitting PDF based on AIC and an appropriate model based on the Anderson-Darling p-value.

Table 14. Navy/EAS Distribution—ESLOC (K)

	Median	Distribution	AD
Navy - EAS	53.56	Lognormal	0.291

The final overlap occurs with Navy/EAS projects. For these projects, a lognormal distribution is the best PDF to use and is also appropriate based on the Anderson-Darling p-value.

ESLOC—Commodity and Contract Type

This analysis investigates the impact on ESLOC as contract type and commodity are changed. Like the Cost/ESLOC sections regarding contract type, since not all data points had viable contract information, some needed to be scrubbed from the dataset. The first analysis holds contract type constant while varying commodity. Three different contract types had detected differences between commodities. Within CPAF, Rotary Wing was found to be smaller than Aircraft. Within CPIF, EAS was found to be larger than Aircraft. Lastly, within MC, UAV was found to be smaller than Space.

Next, commodity was held constant while varying contract type to see the impacts on ESLOC. Within the EAS commodity, CPFF was found to be larger than CPAF, CPIF, and MC. Within the UAV commodity, CPIF was found to be larger than MC.

Overlaps in the two iterations of Steel-Dwass test highlight which specific combinations of commodity and contract type warrant a more specific distribution. Two such overlaps occur, and each distribution is shown in the following tables.

Table 15. EAS/CPFF Distributions—ESLOC (K)

	Median	Distribution	AD
EAS - CPFF	475.791	Exponential	0.052
		Lognormal	0.057

Table 15 reflects the distribution parameters that form a PDF modeling the data for EAS commodities performed with a Cost-Plus Fixed Fee contract. Based on AIC, exponential is the best fitting distribution but lognormal is also provided due to the practitioner's familiarity. Both distributions are appropriate methods to model this data given that both Anderson-Darling p-values are larger than the alpha level of 0.05.

Table 16. UAV/MC Distributions—ESLOC (K)

	Median	Distribution	AD
UAV - MC	25	Gamma	0.8
		Lognormal	0.645

The last overlap occurs for UAV elements performed under a Mixed Contract. Based on AIC, a Gamma distribution was the best fit and both it and a lognormal distribution would be appropriate means to model this specific data since both Anderson-Darling p-values are larger than the alpha level of 0.05. As in previous distributions pertaining to Mixed Contracts, the utility is difficult to define since Mixed Contracts can vary drastically. There is no clear definition of a Mixed Contract Other than possessing fixed and cost-plus elements. The proportions, however, are not defined.

Results, Limitations, and Future Research

This research was oriented toward identifying the various distributions that can be used to model the values of Cost per ESLOC and ESLOC within software programs. Segmenting the dataset by different program characteristics (e.g., service, commodity, contractor, and contract type) highlighted elements of a project that can influence the size and cost of software in defense programs. Additionally, by incrementally changing various characteristics, one can see the marginal changes in each dependent variable as a certain project element is varied.

Results

The findings from this research emphasize the heterogeneity found in Cost per ESLOC and ESLOC values. Although overall distributions can be used to model these values, the results shown earlier indicate that certain characteristics of a project can change the region of plausible values and can aid in creating more specific distributions. The results show that some contractors, commodities, services, and contract types tend to result in bigger or more expensive program elements. Knowing this, it may not always be advisable to use a general distribution when a more specific one is available.

The incremental analyses showed how Cost per ESLOC and ESLOC changed within certain program characteristics. The incremental analyses served the same purpose as a linear regression in that it analyzed how a dependent variable (Cost/ESLOC, ESLOC) varied when another is held constant. Put another way, it showed the marginal changes in the dependent variable because of a change in an independent variable. Each pair of analyses (those with the same independent variables but the one held constant and the one varied were switched) were compared and when overlaps in Steel-Dwass pairs were present, this highlighted the need for a more specific distribution, one tailored to a particular pair of characteristics.

The following flowcharts (see Figure 1 and Figure 2) provide a roadmap for the practitioner to arrive at a recommended distribution to use in their software cost model. The practitioner starts at the left side with the commodity of interest. Moving to the right, one enters another characteristic of the program, in this case it is service. If any of the conditions are met within service, the distribution identifier is provided, and the practitioner stops. (Note: Readers can contact the authors for specific distributions.) If no conditions are met in the service section, the user moves to the next section. Once the user has moved through the entire flowchart, if no intermediate conditions have been met, the identifier for the overall distribution of that commodity should be used. If multiple distributions apply to a given project, any of them can be used to model outcomes; however, it will be at the practitioner's discretion to determine which to use.

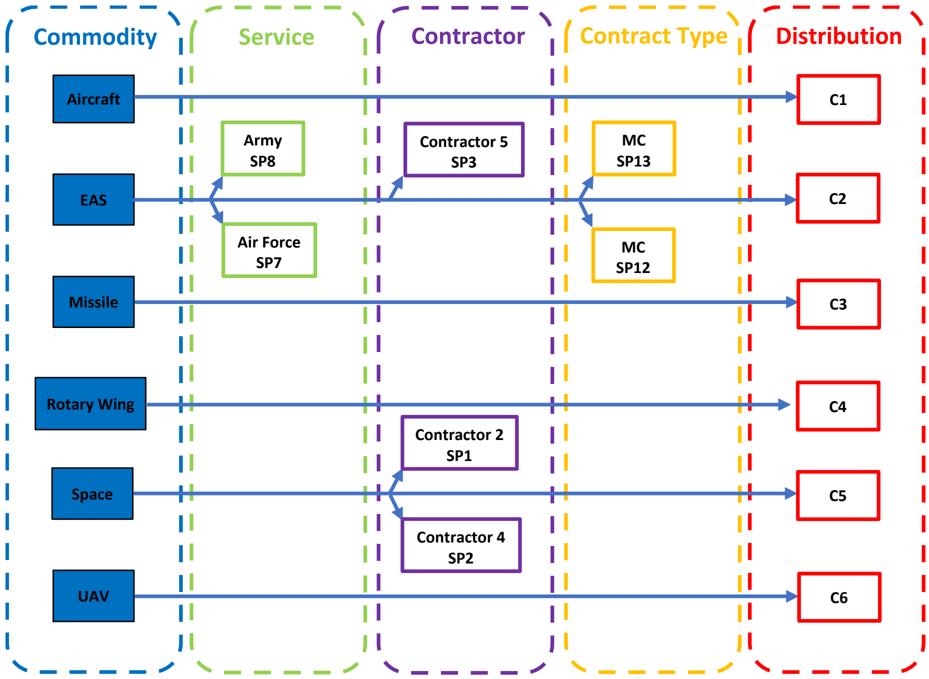

Figure 1. Cost/ESLOC Distribution Flowchart

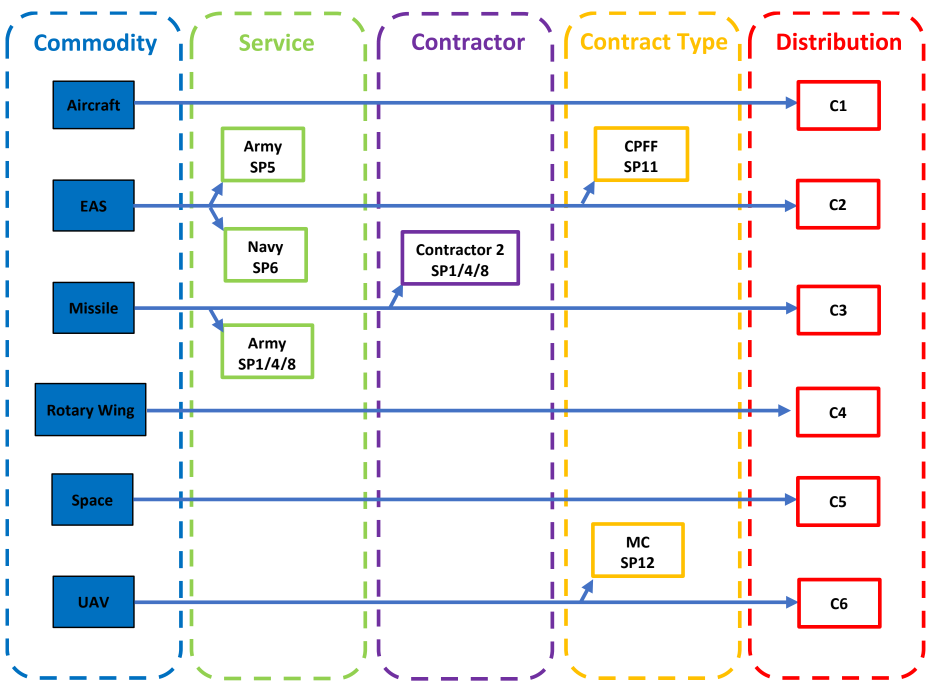

Figure 2. ESLOC Distribution Flowchart

The flowchart could have begun with any of the characteristics; however, commodity was chosen for the following reasons. If this were to be performed very early in the decision-making process, contractor and contract type may not be known. System Type was not used because it is a subset of commodity, subjective in nature, and in many instances, it simply mirrors the results of commodity. Lastly, service was not used since this research was primarily intended for use at AFLCMC, an Air Force entity. For their purposes, they are interested in Air

Force programs and if service was the origin of the flowchart, some specific distributions would be left out, perhaps distributions that would better fit the program of interest.

Limitations

Limitations to this research are mostly related to the dataset. Regarding the various program characteristics, some do not have a well-defined definition and therefore introduce subjectivity. “Mixed contract” does not have a clear definition outlining the proportion of fixed and cost-plus elements. This means that two contracts with wildly different proportions could both be considered a mixed contract and would therefore utilize the same distribution. There were differences found between certain contract types so it would be beneficial to know proportions of fixed and cost-plus elements.

There were also limitations regarding the process used to obtain results. As mentioned before, a traditional regression analysis could not be performed due to overlaps and interaction found in the dataset. For this reason, the incremental approach was taken and although it is a rather laborious substitute, the rationale is largely the same. Since it was not a regression analysis, coefficients were not calculated and thus, firm conclusions regarding a particular characteristic’s impact could not be illustrated, only direction.

Lastly, some projects had more lines in the dataset than others meaning it was more represented in each characteristic. As a result, some commodities, contractors, etc. had more data points not because there were more projects but because there were more WBS elements.

Future Research

If more data can be collected and utilized for these purposes, other methods could be employed in future research. A conventional regression analysis could be performed, and the coefficients would indicate the true impact on Cost per ESLOC and ESLOC. In theory, a regression equation could be formulated to predict the size and cost of a program given only characteristics, a tool that could prove to be invaluable to cost estimators and high-level decision makers alike.

One area that is ripe for future research involves team productivity. One could utilize the cost data in this research and analyze it as a rate of dollars spent per manhour on the project. Also, they could move toward efficiency and investigate hours/ESLOC. Both rates could be analyzed through each characteristic to expose differences and highlight in which situations software development teams tend to be more productive or efficient.

This research illuminates the patterns which costs, and effective sizes follow with regards to various elements of a software program. With these software cost and size distributions, a practitioner can pick the distribution that applies the project they are estimating and know it was created for that exact situation. This research serves as a first step in identifying distributions between software program elements and the costs that are incurred as a result, all with the intent to increase the overall accuracy and effectiveness of cost estimation.

References

Air Force Cost Analysis Agency & Naval Center for Cost Analysis. (2008). *Software development cost estimating handbook.* https://www.dau.edu/cop/ce/dau%20sponsored%20documents/sw%20cost%20est%20manual%20vol%20i%20rev%2010.pdf

Clark, B., & Madachy, R. (2015). Software cost estimation metrics manual for defense systems. *Software Metrics Inc.*

Garrett, G.A. (2008). Cost estimating methods and Sources of risk. *Contract Management, 48*(10), 14–24.

Jones, G., White, E., Ryan, E. T., & Ritschel, J. D. (2014). *Investigation into the ratio of operating and support costs to life-cycle costs for DoD weapon systems*. Air Force Institute of Technology Wright-Patterson AFB OH.

Keung, J. (2009). Software development cost estimation using analogy: A review. *2009 Australian Software Engineering Conference.* https://doi.org/10.1109/aswec.2009.32

Klimberg, R., & McCullough, B. D. (2016). *Fundamentals of predictive analysis with Jmp* (2nd ed.). SAS Institute Inc.

McQuade, J. M., Murray, R. M., Louie, G., Medin, M., Pahlka, J., & Stephens, T. (2019, March 12). *Periodic table of Devops tools V3 PDF*. Retrieved July 28, 2021, from https://media.defense.gov/2019/Mar/14/2002101480/-1/-1/0/DIB-SWAP_STUDY_REPORT[DRAFT],_LAST%20MODIFIED_13MAR2019.PDF.

Pfleeger, S. L., Wu, F., & Lewis, R. (2005). *Software cost estimation and sizing methods: Issues, and guidelines.* Rand Corp.

Shepperd, M., & Schofield, C. (1997). Estimating software project effort using analogies. *IEEE Transactions on Software Engineering, 23*(11), 736–743. https://doi.org/10.1109/32.637387

Acquisition Research Program
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net