Exploring DoD Software Growth: A Better Way to Model Future Software Uncertainty

Primary Authors:
Nicholas Lanham
Aviation Estimating Division
Naval Center for Cost Analysis

Dr. Corinne Wallshein
IT Estimating Division
Naval Center for Cost Analysis

Acknowledgements:
Dr. Wilson Rosa
IT Estimating Division
Naval Center for Cost Analysis

Mike Popp
Software & Acquisition Estimating
NAVAIR 4.2, OMNITEC

ABSTRACT

Derived from the April 2014 Software Resource Data Report (SRDR) dataset, this analysis provides the Department of Defense (DoD) cost estimating community with new estimating relationships to predict software in development. These relationships are based on Initial variables, singularly and combined. Innovative metadata tags facilitate higher quality software growth uncertainty modeling. As a result of updating and revising a paired data algorithm in the SRDR table posted to Defense Cost And Research Center's (DCARC) Defense Automated Cost Information Management System (DACIMS) web portal, several new variables were added for lower-level analysis. New Contract Type, Application Domain (AD), Super Domain (SD), Program Type, and Operating Environment (OE) variables enable different subdivisions and analysis of SRDRs. All of the following relationships were derived from multiple Department of the Navy (DON) cost organizations' review of 438 Computer Software Configuration Items (CSCIs) paired to analyze a dataset of 219 DCARC-approved SRDR data points to reflect DoD-specific software development efforts.

KEYWORDS

Software cost estimation, software growth, application domain, super domain, contract type, final hours, multivariate software relationships.

1.0 INTRODUCTION

1.1 Problem Statement

Prior studies have attempted to predict software development hours and potential software growth by using Delphi techniques with Subject Matter Expertise inputs. Most lack lower-level analysis of actual DoD-specific software development trends. While these studies have been used to estimate a multitude of new, software-intensive weapon systems, the DoD cost estimating community continues to make significant updates to the data collection processes, analysis, and parsing strategies specific to DoD SRDR data.

Utilizing improved data collection, verification, validation, and parsing processes, this study leverages DoD-specific data to analyze the percent change in software development hours by Contract Type, SD, and AD to assist the cost estimating community with better predictive relationships in each of these environments. This study also highlights improved accuracy of predicting Final software development hours from initial hour estimates, vice inconsistent software size-based estimates such as Equivalent Source Line of Code (ESLOC) productivity (Hours per Source Line of Code) metrics.

1.2 Purpose of Study

This study expands existing DoD software cost estimating knowledge base by analyzing the influence of Contract Type, AD, and SD on the percent change from initial to final software reporting events. The percent-change analysis of total development hours is one focus of this report. Future analysis may exploit Mr. Nicholas Lanham's innovative SRDR pairing algorithm, included in the latest SRDR dataset posted to the DACIMs web portal, which allows analysts to compare numerous other variables percent change from initial to final reporting events (e.g., requirements counts, peak staffing, SLOC, ESLOC, hours by software development phase, and requirements volatility)

A second analysis focus area provides the DOD cost estimating community with two effort estimating relationships to better predict Final software development hours using estimates of Initial development hours, Initial count of total requirements, and/or Initial count of peak staff.

2.0 Related Work

As part of the recent Air Force Cost Analysis Agency (AFCAA)-led SRDR Working Group (SRDRWG), the Naval Center for Cost Analysis (NCCA) analyzed each CSCI in the latest Naval Air System Command Cost Department (NAVAIR 4.2)-developed SRDR dataset, as of April 2014. Mr. Lanham expanded on an existing CSCI Pairing Algorithm; he processed each CSCI submission and grouped "Pairs" by matching an Initial (2630-2) record with its corresponding Final (2630-3) record. The resulting dataset has 219 paired records. These records also passed a detailed quality screening process during data entry.

Questions and guidelines used to complete this quality screening process were outlined in Mr. Lanham and Mr. Popp's SRDR Verification and Validation (V&V) guide, currently in its final draft version. It includes specific questions and checklists to aid analysts performing a quality assessment of SRDR data records to be included in the final dataset. This SRDR V&V guide also includes AD, SD, and OE definitions developed by Dr. Wilson Rosa, Cheryl Jones, et al[1].

[1] SRDRWG Data Collection Brief on CSCIs by Application Domain (2014)

Current SD, AD, and OE definitions are in the coordination version of the update to the SRDR Data Item Description (DID). Organizations submitting SRDR documentation must provide SD, AD, and OE designations for each CSCI in their report(s). In the future, NAVAIR 4.2 and SRDR Unified Review Function (SURF) team will reference an official, final SRDR V&V guide to ensure the submitting organization(s) made defensible AD, SD, and OE designations for each CSCI during SRDR submission.

Since the definitions are being finalized as part of the ongoing SRDRWG discussion forum, the original NAVAIR 4.2-developed dataset did not include the additional AD, SD, and OE data tags. Afterwards, a team of NCCA analysts assessed each individual SRDR data submission and added the most appropriate AD, SD, and OE data tag per a consistent, early version of the SRDR V&V guide.

3.0 RESEARCH METHOD

3.1 Research Questions

This study will address the following questions:

What SRDR variables exhibit similar percent change trends from initial to final reporting events?

Does Contract Type influence the percent-change (i.e., growth) in total software development hours from Initial to Final reporting events?

Is there another method to accurately predict Final software development hours from an Initial estimate, other than ESLOC conversions and productivity rate metrics?

The following data collection and analysis methods may enable analysts to obtain a deeper understanding of the underlying trends and relationships driving modern DoD software development efforts. Seven-plus years of data collection and analysis are incorporated in the

April 2014 NAVAIR-developed SRDR dataset containing every CSCI submitted by industry to DCARC since 2004.

4.0 PERCENT CHANGE TRENDS

As part of a recent update to the SRDR pairing algorithm, existing and new software variables were designated to isolate growth trends in DoD software development efforts. Software growth distributions were derived for these variables; a portion is described in Appendix A.

Figure 1 illustrates statistics specific to several key software development variables to assist the cost community to more effectively isolate data-driven growth trends for DoD weapon systems. In analyzing levels of change from the submitting organization's initial and final reporting events using raw SRDR data submissions, Figure 1 identifies central tendencies in growth by size, effort hours, duration, and requirements showing the average or mean value and the median. Variability is characterized by standard deviation (Std Dev.) and the coefficient of variation (CV).

Percent change was derived using the following formula and was applied to the each of the records included in the Paired Dataset.

$$\frac{(Final\,Var.X - Initial\,Var.X)}{Initial\,Var.X} = \%\,Change\,in\,Var.X$$

NCCA identified the percent change from initial to final reporting events specific for variables of total SLOC, new SLOC, requirements count, development hours, and development duration. The percent change behaved differently for each software development variable in Figure 1. This analysis identifies software growth trends and variability specific to several key SRDR variables in this 219 point paired data set. At this point in time, this analysis should assist cost analysts to better reflect historical DoD software growth trends as they develop software cost and uncertainty models in the near term and help guide them in analyzing future data.

SRDR Variable Percent Change	**Mean**	**Median**	**Std Dev.**	**CV**
Total SLOC	1.33	0.05	10.66	799.15
New SLOC	1.6	0.29	7.51	468.51
Total Dev. Hours	0.78	0.23	1.77	226.43
Total Dur. (Months)	0.88	0.08	5.69	647.26
Total Req. Count	3.62	0.00	35.1	969.35

Figure 1: Percent Change From Initial to Final Reporting Events by Variable

This analysis generated differing mean percent change values for each key variable in Figure 1. The percent change for the total requirements count had the lowest median value of 0.0 while exhibiting the highest variability. Percent change for total SLOC and new SLOC exhibited mean values between 1.3 and 1.6 and median values below 0.3. Since total SLOC sums up all categories of SLOC, the standard deviation for total SLOC was expectedly higher than the standard deviation for new SLOC. The highest level of variability was for the total requirements count, followed by the next highest level of variability for total software size in SLOC.

Another interesting relationship was identified between total development hours and total duration in months. The variables had similar mean values as total development hours averaged 0.78 and total duration averaged 0.88, median values diverged from 0.23 for hours to 0.08 for months. Variability for total duration was the third highest in Figure 1, whereas variability in total hours was the lowest. Schedule may stretch as requirements are culled, possibly leading to refinements in software size estimates.

5.0 CONTRACT TYPE EFFECTS

Contract numbers in SRDR submissions were used to research contract types. Contract types were added as a variable to provide the DoD cost estimating community with a novel way to study software growth. This variable was used to determine how, or whether, Contract Type influences growth on historical DoD software development efforts. Using contract numbers to determine the resulting contract type, NCCA cross-referenced several common DoD databases, such as Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD/AT&L) Defense Acquisition Management Information Retrieval (DAMIR), DCARC Cost Assessment Data Enterprise (CADE), and Defense Logistics Agency (DLA) Electronic Document Access (EDA), and online sources. This categorical metadata tag was used to analyze the percent-change in total hours by each contract type in the dataset.

After sorting by contract type, the overwhelming majority of the paired data were Cost Plus Incentive Fee (CPIF), Cost Plus Award Fee (CPAF), or Cost Plus Fixed Fee (CPFF) contracts. A small portion of Fixed Priced (FP) and Indefinite Delivery Indefinite Quantity (IDIQ) contracts were included in the final paired dataset. These contract types lacked sufficient quantities of records to adequately compare to Cost Plus Contract Types.

Percent Change in software development hours was derived using the following formula on each record in the paired Dataset:

$$\frac{(Final\ Hours - Initial\ Hours)}{Initial\ Hours} = \%\ Change\ in\ Hours$$

Figure 2 illustrates the dispersion of the percent change (e.g., growth) in total hours from initial to final reporting events by contract type.

Level	Number	Mean	Std Dev.	CV
CPAF	91	1.12232	2.1379	190.49
CPFF	50	0.38451	0.7935	206.36
CPIF	43	0.79150	2.0324	256.78

Figure 2: Contract Type ANOVA Statistics

Using the SAS JMP Statistical Discovery™ software package, the means of each Cost Plus contract type were compared using a Student's T-Test. Before making this comparison, the analysis team did not adjust, modify, or remove data. Analysis includes raw, unadjusted values for total development hours. At first glance, CPAF contracts include a substantially larger mean percent-change factor of 112% compared to CPFF and CPIF. Figure 3 further highlights variance associated with each Cost Plus contract type record.

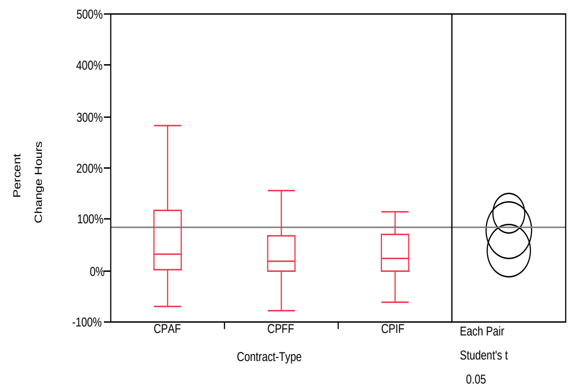


Figure 3: Contract Type Variance Analysis, zoomed view from -100% to 500%

Using the Student's T Pairwise comparison of each combination of Cost Plus contract types, CPAF and CPFF resulted in means that represented statistically significant differences. This difference highlights a need to better understand performance drivers by contract type from initial to final reporting events. CPAF contracts are typically structured to pay a specific fee at the time of award and provide the potential for additional fees based upon the contractor's performance in key areas. While this analysis does not attempt to assess the final performance of any given system, the resulting variation in the percent-change of total hours specific to CPAF contract type highlights the DoD contracting community's historical challenge to independently and accurately estimate final software development hours.

DoD cost analysts may now point to empirical evidence supporting different software growth relationships for each Cost Plus contract type. Means and standard deviations may be used to inform uncertainty cost estimation parameters when applying adjustments to Initial development hours, if contract type is known at the time of estimation. This investigation into how reported software development hours change from initial to final reporting events has generated a series of new questions. When assessing software development growth rates, why do CPAF contracts behave significantly different from CPIF and CPFF contracts? Why are standard deviations for CPAF and CPIF similar? Why does CPFF have such a relatively low standard deviation? These are questions for the DoD cost analysis community to ask as they develop new software development estimates.

Contract type for the 219 paired records appears to impact mean and standard deviation values for software growth in historical DoD software development efforts meeting or exceeding DCARC SRDR cost reporting thresholds (i.e., contract values over $20M). The answer to the first research question is yes. Contract type influences the growth, measured in percent-change, in total development hours from initial to final reporting events.

6.0 INITIAL AND FINAL HOURS

Due to various, discrete methods of converting SLOC into ESLOC, the DoD cost estimating community encounters scenarios where Hours per ESLOC metrics have been derived and implemented inconsistently. When this happens, the ratio of Hours per ESLOC varies, making estimates of new software efforts incomparable to historical efforts. ESLOC figures across CSCIs, programs, or other variable subsets are software size measures. If the underlying ESLOC calculation formulas are not provided, software size in ESLOC and productivity measured in Hours per ESLOC, frequently used in the cost community, are obscured. ESLOC measures do not provide an optimal way to compare software size. Hours per ESLOC productivity metrics do not provide an optimal methodology to predict final software development hours. When a cost analyst is provided an initial estimate of development hours from a contracting or other software development organization in ESLOC, the first question should be "what are the underlying parameters and what software measurement data was used as inputs?"

To provide the cost estimating community with another method of estimating software development effort without relying on ESLOC conversions and/or Hours per ESLOC productivity metrics, the 2014 paired SRDR database provided a basis to derive effort and schedule estimating relationships.

Variable	F. Hours	I. New	I. Mod	I. Reuse	I. Auto	I. Hrs.	I. Req.	I. Peak Staff	I. Month
F.Hours	1.00	0.59	0.27	0.27	0.11	0.89	0.40	0.78	0.07
I.New	0.59	1.00	0.22	0.20	0.03	0.64	0.20	0.49	0.13
I.Mod	0.27	0.22	1.00	0.08	0.05	0.20	0.59	0.10	0.07
I.Reuse	0.27	0.20	0.08	1.00	0.13	0.28	0.13	0.20	0.07
I.Auto	0.11	0.03	0.05	0.13	1.00	0.07	0.06	0.06	0.00
I.Hrs.	0.89	0.64	0.20	0.28	0.07	1.00	0.32	0.72	0.09
I.Req	0.40	0.20	0.59	0.13	0.06	0.32	1.00	0.25	0.14
I.Peak Staff	0.78	0.49	0.10	0.20	0.06	0.72	0.25	1.00	0.01
I.Month	0.07	0.13	0.07	0.07	0.00	0.09	0.14	0.01	1.00

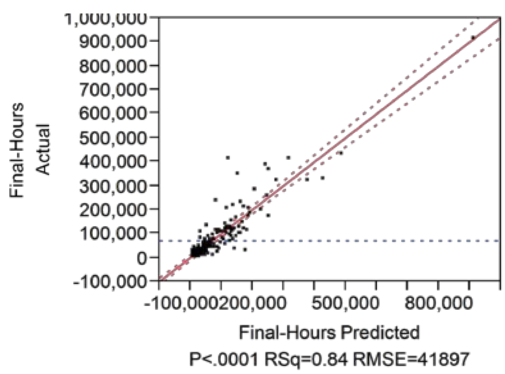
Figure 4: Initial to Final SRDR Variable(s) Correlation Matrix, REML Method

The correlation matrix shown in Figure 4 was initially developed to understand what Initial variables are related to the Final Hours. Initial and Final Months (Duration) for each SRDR record were derived from taking the minimum start date and maximum end date reported for each IEEE 12207 phase and calculating the delta between them to arrive at total duration, in months. In the dataset, Final hours was highly correlated to initial hours and initial peak staffing – not to SLOC and/or ESLOC.

Linear regression equations were developed to estimate Final hours by Initial variables. The following equation was derived using a linear, multivariate regression model with Final hours as the dependent variable and Initial hours and requirement counts as the independent variables. Mr. Lanham revised the paired data algorithm to include additional Initial SRDR variables within the pairing routines. This update provides analysts with the capability to perform in-depth explorations of growth and software trends.

Subsequent formulas, as well as the supporting appendices, allow analysts to generate estimates for Final hours using Initial hours, Initial requirements count, and/or Initial peak staffing vice using ESLOC conversions and productivity metrics.

$$Estimated\ Final\ Hours = 834.46 + .871 * Initial\ Hours + 10.520 * Inital\ Req.Count + 824.215 * Initial\ Peak\ Staff$$



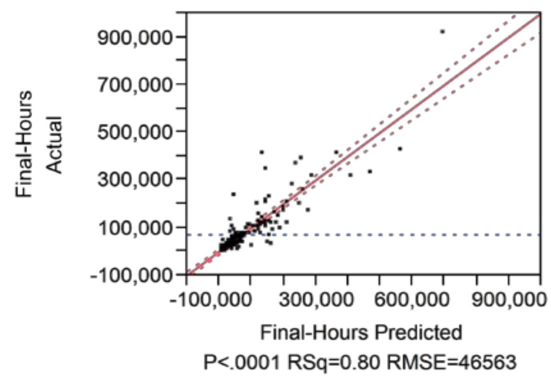
Summary of Fit Statistics	Value:
RSquare	0.837
RSquare Adj	0.834
Root Mean Square Error	41,897.34
Mean of Response	70,323.35
Observations (or Sum Wgts)	214

Figure 5: Final Hours by Initial Hours, Peak Staffing, and Requirements Count

The regression equation illustrated in Figure 5 use Initial hours, Initial requirements count, and Initial peak staffing as inputs. Although this equation exhibits very strong fit statistics, the study team continued to probe relationships between Final hours, Initial hours, and Initial requirement counts to develop an equation with these few, named variables. As shown by Figure 4, Initial hours is correlated at 0.72 with Initial peak staffing.

Figure 6 illustrates a multivariate linear equation requiring an estimate of Initial Hours and Initial Requirement Counts.

$$Estimated\ Final\ Hours = 2{,}013.37 + 1.151 * Initial\ Hours + 11.162 * Inital\ Req.Count$$



Summary of Fit Statistics	Value:
RSquare	0.798
RSquare Adj	0.796
Root Mean Square Error	46,563.46
Mean of Response	70,323.35
Observations	214

Figure 6: Final Hours by Initial Hours and Requirement Counts

DoD cost analysts may use Initial hours and Initial requirement counts to accurately estimate Final hours by leveraging trends and relationships in over 200 raw, actual SRDR CSCI data submissions. These effort estimating equations and their supporting statistics provide realistic and accurate methods to crosscheck software development effort estimates when using other models, tools, or equations.

These formulas also provide cost analysts with multiple alternatives to accurately estimate Final hours based on given Initial variables vice being reliant upon inconsistent ESLOC conversion factors and/or Hours per ESLOC productivity metrics. Relationships described in this study are intended to aid your cost analysis team(s) as they develop their next software cost estimate. The answer to the second research question is yes. There are at least two other methods to accurately predict Final software development hours from

Initial estimates of hours, requirement counts, and/or peak staffing.

7.0 CONCLUSION

After manual efforts to research and record the Contract Type and choose the most appropriate AD and SD and Mr. Lanham's development of a SRDR pairing algorithm to match Final CSCI records to Initial CSCI records, this analysis highlighted the different Contract Type percent-change or growth in software development hours by CPAF, CPIF, and CPFF.

With the entire, unaltered set of paired records, two effort estimating equations using Initial hours, Initial requirements count, and/or Initial peak staffing were found to accurately and adequately predict Final hours. Final hours may be estimated using equations in Figure 5 or Figure 6.

8.0 FUTURE WORK

To further expand the software growth analysis and regression relationships discussed in this paper, NCCA will continue to analyze the percent change specific to total SLOC, new SLOC, development hours, duration, and requirements count by identified subsets such as super domain, application domain, and operating environment. As a result of NCCA's recent metadata tagging efforts, the existing SRDR dataset can be filtered to derive additional multivariate regression equations by super domain, application domain, and operating environment to provide the cost community with a more robust set of estimating relationships to derive final development hours using estimates of initial hours and/or total requirement counts.

Continued analysis on this dataset and comparative analysis on subsequent dataset updates will allow experimentation on specific data subsets. Subsets may use inherent categorizations such as new versus update, Capability Maturity Model Integrated (CMMI) levels, and platform. As the SURF team is activated and routinely performs quality data assessments, future data will be scrutinized more thoroughly by the cost community, improving the number of records considered to be "good" for analysis. Future analysis, including new categorizations of effort by phase, may uncover consistent effort estimating relationships and percent-change relationships.

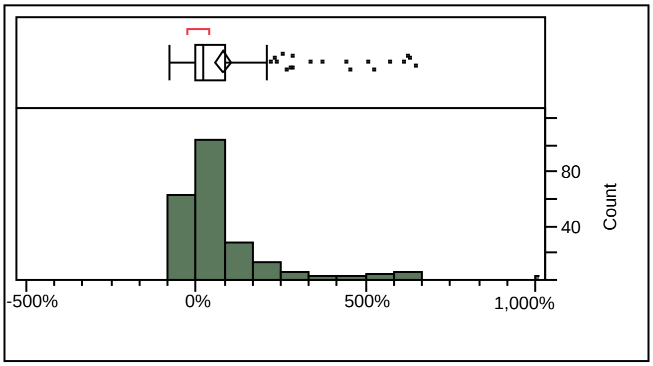
Appendix A – Percent Change Analysis

This appendix has distributions of the percent change from Initial (2630-2) to Final (2630-3) SRDR reports. Each paired data point is evaluated at the lowest CWBS/CSCI to ensure that the initial and iinal data submissions comply with the latest version of the DoD SRDR V&V guide. After confirming data compliance and data quality, the resulting paired dataset includes an algorithm to compare the percent change of several critical SRDR variables. This appendix provides a statistical overview of each of the analyzed variables and illustrates how the variable has changed from initial to final reporting events.

The following table provides a summary of the supporting ANOVA statistics specific to the percent change analysis specific to total development hours, total SLOC, total requirements count, total development duration, and new SLOC. In addition, the detailed statistics and distribution characteristics have been provided within the following pages.

Percent Change	**Total Development Hours**	**Total SLOC**	**Total Requirements Count**	**Total Duration (Development Months)**	**Total New SLOC**
Mean	0.78	1.33	3.62	0.88	1.60
Std Deviation	1.77	10.66	35.10	5.69	7.51
Skewness	3.68	13.85	12.25	11.17	11.17
Kurtosis	16.42	199.15	158.01	137.31	142.16
CV	226.43	799.15	969.35	647.26	468.51
Median	0.23	0.05	0.00	0.08	0.29

Percent Change in Total Development Hours



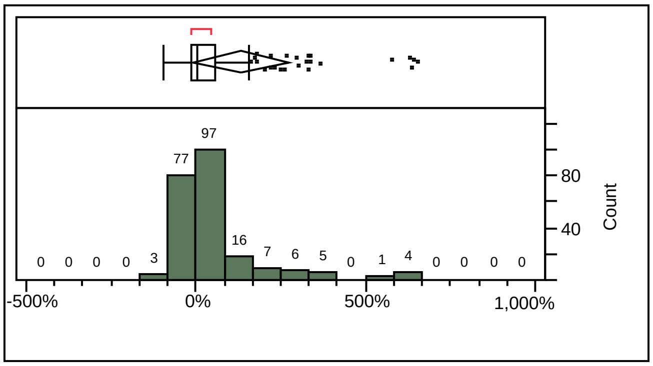
Quantiles

100.0%	maximum	11.6154
99.5%		11.5737
97.5%		6.26613
90.0%		2.22239
75.0%	quartile	0.85239
50.0%	median	0.2278
25.0%	quartile	-0.0132
10.0%		-0.2762
2.5%		-0.6444
0.5%		-0.7783
0.0%	minimum	-0.7786

Moments

Mean	0.7835791
Std Dev	1.7742398
Std Err Mean	0.119892
Upper 95% Mean	1.019875
Lower 95% Mean	0.5472833
N	219
Sum Wgt	219
Sum	171.60383
Variance	3.1479269
Skewness	3.682413
Kurtosis	16.417546
CV	226.42765
N Missing	0

Percent Change in Total SLOC



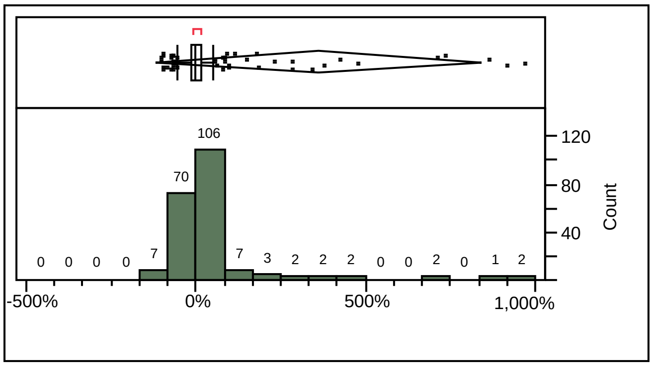
Quantiles

100.0%	maximum	154.784
99.5%		141.538
97.5%		6.39031
90.0%		2.1893
75.0%	quartile	0.5615
50.0%	median	0.04695
25.0%	quartile	-0.115
10.0%		-0.5225
2.5%		-0.8083
0.5%		-0.9448
0.0%	minimum	-0.9468

Moments

Mean	1.3345232
Std Dev	10.664893
Std Err Mean	0.7206667
Upper 95% Mean	2.7548892
Lower 95% Mean	-0.085843
N	219
Sum Wgt	219
Sum	292.26058
Variance	113.73994
Skewness	13.84507
Kurtosis	199.14839
CV	799.15381
N Missing	0

Percent Change in Total Requirement Counts



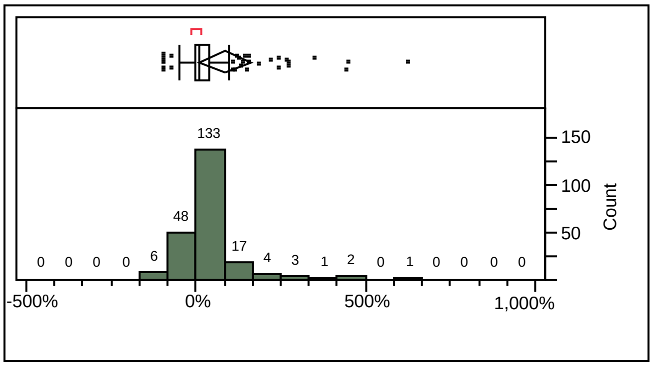
Quantiles

100.0%	maximum	472.389
99.5%		458.119
97.5%		11.1187
90.0%		1.13523
75.0%	quartile	0.14197
50.0%	median	0
25.0%	quartile	-0.1259
10.0%		-0.5636
2.5%		-0.9776
0.5%		-1
0.0%	minimum	-1

Moments

Mean	3.6212649
Std Dev	35.102815
Std Err Mean	2.4281125
Upper 95% Mean	8.40813
Lower 95% Mean	-1.1656
N	209
Sum Wgt	209
Sum	756.84437
Variance	1232.2076
Skewness	12.251275
Kurtosis	158.01385
CV	969.3523
N Missing	10

Percent Change in Total Duration (Development Months)



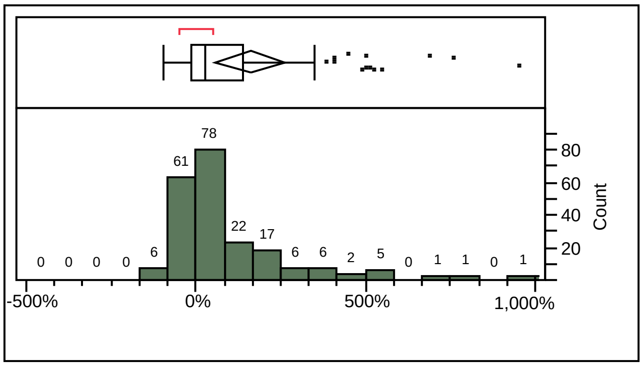
Quantiles

100.0%	maximum	74.8333
99.5%		70.8242
97.5%		4.44376
90.0%		1.21484
75.0%	quartile	0.40122
50.0%	median	0.08413
25.0%	quartile	-0.0025
10.0%		-0.1843
2.5%		-0.9369
0.5%		-0.9799
0.0%	minimum	-0.9817

Moments

Mean	0.878927
Std Dev	5.6889451
Std Err Mean	0.385304
Upper 95% Mean	1.6383443
Lower 95% Mean	0.1195097
N	218
Sum Wgt	218
Sum	191.60608
Variance	32.364096
Skewness	11.172894
Kurtosis	137.31376
CV	647.26026
N Missing	1

Percent Change in Total New SLOC



Quantiles

100.0%	maximum	100
99.5%		95.6362
97.5%		10.0153
90.0%		3.42067
75.0%	quartile	1.40604
50.0%	median	0.29354
25.0%	quartile	-0.1587
10.0%		-0.5207
2.5%		-0.8767
0.5%		-0.9399
0.0%	minimum	-0.9401

Moments

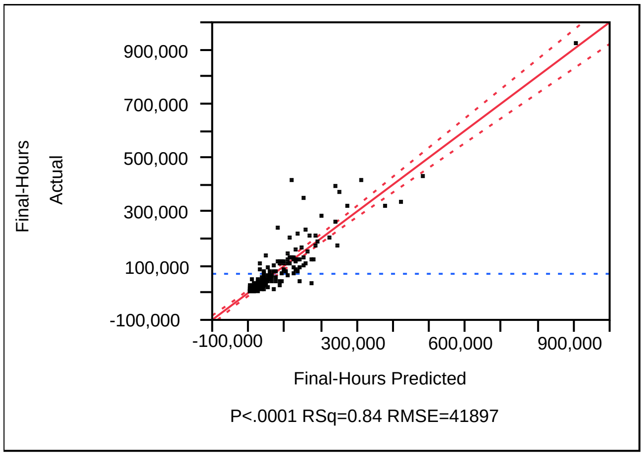
Mean	1.6039412
Std Dev	7.5145702
Std Err Mean	0.5173243
Upper 95% Mean	2.6237555
Lower 95% Mean	0.5841269
N	211
Sum Wgt	211
Sum	338.43159
Variance	56.468765
Skewness	11.17279
Kurtosis	142.15792
CV	468.50659
N Missing	8

Appendix B – Predicting Final Hours with Initial Variables, Includes all Paired SRDR Data

This appendix contains a new approach to predict final hours based on a multivariate linear regression model. This dataset includes 100% raw SRDR data and illustrates a multivariate estimating model based upon an optimal mix of independent variables analyzed within the paired dataset. After analyzing final hours related to each provided initial SRDR variable, the following multivariate model had the highest adjusted R-squared and lowest Root Mean Squared Error (RMSE) to estimate final hours. This appendix has a statistical summary of the following multivariate model used to predict final hours, with initial hours, initial requirement counts, and initial peak staffing as independent variables.

Final Hours = 834.46 + 0.87 * Initial Hours + 10.52 * Initial Requirements + 824.22 * Initial Peak Staffing

Final-Hours by Initial Hours, Requirement Counts, and Peak Staff Includes All Data



Summary of Fit

RSquare	0.837236
RSquare Adj	0.83491
Root Mean Square Error	41897.34
Mean of Response	70323.35
Observations (or Sum Wgts)	214

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	3	1.8962e+12	6.321e+11	360.0697
Error	210	3.6863e+11	1.7554e+9	**Prob > F**
C. Total	213	2.2648e+12		<.0001*

Lack Of Fit

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Lack Of Fit	209	3.6863e+11	1.7638e+9	.
Pure Error	1	0	0	**Prob > F**
Total Error	210	3.6863e+11		.
				Max RSq
				1.0000

Parameter Estimates

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	834.46099	3642.273	0.23	0.8190
Initial-Hours	0.8716524	0.056174	15.52	<.0001*
Initial Req. Count	10.519638	2.549626	4.13	<.0001*
Inital-Peak-Staff	824.2153	115.851	7.11	<.0001*

Prediction Expression

834.460986758451

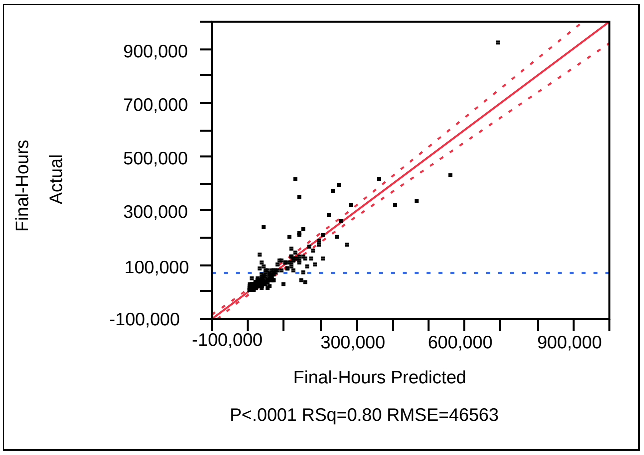
+ 0.87165242878067 * Initial-Hours
+ 10.5196378957462 * Inital-Req
+ 824.215301198734 * Inital-Peak-Staff

Appendix C – Predicting Final Hours with Initial Variables, Includes All Paired SRDR data

This appendix includes a new approach to predicting final hours based on a multivariate linear regression model with two independent variables. Based on a dataset of 100% raw SRDR data, a multivariate estimating model uses an optimal mix of independent variables analyzed in the paired dataset. This appendix includes a statistical summary of a multivariate model used to predict final hours, which includes initial hours and initial requirement counts as independent variables. The statistics for this effort estimating relationship follows the equation shown below.

Final Hours = 2013.38 + 1.15 * Initial Hours + 11.16 * Initial Requirements Count

Final-Hours by Initial Hours and Requirement Count
Includes All Data



Summary of Fit

RSquare	0.798005
RSquare Adj	0.796091
Root Mean Square Error	46563.46
Mean of Response	70323.35
Observations (or Sum Wgts)	214

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	1.8073e+12	9.037e+11	416.7912
Error	211	4.5748e+11	2.1682e+9	**Prob > F**
C. Total	213	2.2648e+12		<.0001*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	210	4.5748e+11	2.1785e+9	.
Pure Error	1	0	0	**Prob > F**
Total Error	211	4.5748e+11		.
				Max RSq
				1.0000

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2013.377	4043.722	0.50	0.6191
Initial-Hours	1.1511816	0.044618	25.80	<.0001*
Initial Req. Count	11.162048	2.8318	3.94	0.0001*

Prediction Expression

2013.37701109894

+ 1.15118155052649 * Initial-Hours

+ 11.1620475346908 * Inital-Req