

Modeling median will-cost estimates for defense acquisition programs

Modeling
median will-
cost estimates

19

Ryan Trudelle and Edward D. White

*Department of Mathematics and Statistics, Air Force Institute of Technology,
Wright-Patterson AFB, Ohio, USA, and*

Dan Ritschel, Clay Koschnick and Brandon Lucas

*Department of Systems Engineering and Management,
Air Force Institute of Technology, Wright-Patterson AFB, Ohio, USA*

Received 25 April 2017

Revised 20 June 2017

Accepted 27 June 2017

Abstract

Purpose – The introduction of “should cost” in 2011 required all Major Defense Acquisition Programs (MDAP) to create efficiencies and improvements to reduce a program’s “will-cost” estimate. Realistic “will-cost” estimates are a necessary condition for the “should cost” analysis to be effectively implemented. Owing to the inherent difficulties in establishing a program’s will-cost estimate, this paper aims to propose a new model to infuse realism into this estimate.

Design/methodology/approach – Using historical data from 73 Departments of Defense programs as recorded in the selected acquisition reports (SARs), the analysis uses mixed stepwise regression to predict a program’s cost from Milestone B (MS B) to initial operational capability (IOC).

Findings – The presented model explains 83 per cent of the variation in the program acquisition cost. Significant predictor variables include: projected duration (months from MS B to IOC); the amount of research development test and evaluation (RDT&E) funding spent at the start of MS B; whether the program is considered a fixed-wing aircraft; whether a program is considered an electronic system program; whether a program is considered ACAT I at MS B; and the program size relative to the total program’s projected acquisition costs at MS B.

Originality/value – The model supports the “will-cost and should-cost” requirement levied in 2011 by providing an objective and defensible cost for what a program should actually cost based on what has been achieved in the past. A quality will-cost estimate provides a starting point for program managers to examine processes and find efficiencies that lead to reduced program costs.

Keywords Acquisition estimates, Regression modeling, Should-cost, Will-cost

Paper type Research paper

Introduction

On June 15, 2011, the Under Secretary of Defense for Acquisition, Technology and Logistics [USD (AT&L)] directed the Military Departments and Directors of Defense Agencies via a memorandum to implement will-cost and should-cost management for Acquisition Category (ACAT) I, II and III programs. In this memorandum, the USD (AT&L) reiterates that the departments will continue to set program budget baselines



© In accordance with section 105 of the US Copyright Act, this work has been produced by a US government employee and shall be considered a public domain work, as copyright protection is not available. Published in *Journal of Defense Analytics and Logistics*. Published by Emerald Publishing Limited.

*Journal of Defense Analytics and
Logistics*
Vol. 1 No. 1, 2017
pp. 19-33
Emerald Publishing Limited
2399-6439
DOI 10.1108/JDAL-04-2017-0004

using non-advocate will-cost estimates. A will-cost estimate uses traditional cost-estimating techniques (e.g. analogy, bottom-up, parametric, etc.) to estimate the most likely cost of a program to establish a reasonable budget baseline and acquisition program thresholds. However, the USD (AT&L) also “challenges program managers to drive productivity improvements into their programs during [...] program execution by conducting Should-Cost Analysis”, which involves, “identifying and eliminating process inefficiencies and embracing cost savings opportunities” (Carter and Mueller, 2011). The should-cost estimate therefore deviates below the will-cost estimate to develop a realistic price objective for negotiation purposes and subsequent savings against the will-cost estimate.

Additionally, the USD (AT&L) states in the same memorandum that:

[...] the main problem with the will-cost estimate isn't in the numbers or how it was reached; the problem is that once the will-cost estimate is derived and the budget for the program is set, historically, this figure becomes the “floor” from which costs escalate, rather than a “ceiling” below which costs are contained—in many ways creating a self-fulfilling prophecy of budgetary excess (Carter and Mueller, 2011).

We suggest that perhaps there is a better way to infuse realism into a will-cost estimate such that it becomes a middle-of-the-road estimate from which to work from in the should-cost approach rather than the floor.

Therein lies the crux of the problem – how does one go about generating a median “will-cost” estimate? Defense acquisition programs expand the frontiers of today's technology to develop new and innovative systems that provide an asymmetric advantage on the battlefield. As a result, there are inherent uncertainties and risks associated with Department of Defense (DoD) acquisitions. These realities are manifested in the derivation of the program's cost estimate. To combat risk and uncertainty, cost analysts account for the distribution of possible costs for a program having a right-skewed distribution by estimating to the mean. However, the mode or the most likely cost is less than the mean in skewed distributions. This difference might tie-up resources that may be better placed elsewhere. By contrast, building an overly aggressive cost estimate may free up resources to be placed elsewhere. However, if this estimate is exceeded, decision-makers could take critical funding from other programs or force a program manager to delay the program until additional funding can be secured.

To combat these issues, programs should strive for a realistic, middle ground point – essentially an empirically validated cost baseline. The use of historical data allows the acquisition community to unbiasedly analyze and estimate what a program would cost in relation to other similarly completed programs. This estimate then becomes a powerful tool from which the user can identify a target cost for a given program. This estimate also serves as a benchmark to identify whether a cost estimate is reasonable given what has occurred in the past. From this estimate, mitigation of risks associated with over- and under-estimating program costs may be achieved, resulting in a more efficient allocation of resources. Thus, we propose a new empirically based model for determining will-cost estimates in DoD acquisition programs.

Past research and database creation

Our research builds an empirically derived model to predict median will-cost estimates for DoD acquisition programs. We use prior research to identify potential explanatory variables in our model and establish the basis for creating our data set. Our literature review spans the change in acquisition taxonomy from Milestone (MS)

I, II and III to MS A, B and C. For this study, we consider MS I, II and III to be equivalent to MS A, B and C, respectively. This is consistent with prior literature findings that the naming convention has simply altered over time without tangible changes in definition or substance (Harmon, 2012; Jimenez *et al.*, 2016). Prior to relaying our data collection process, we first discuss recent studies pertinent to our research. These studies provide the foundation for how we conduct our research into the relative program characteristics that predict program cost.

Jimenez *et al.* (2016) developed a schedule duration prediction model for defense acquisition programs using pre-MS B data; we leverage their research to identify explanatory variables for investigation and an initial data set from which to draw upon. Their analysis concluded that the following variables were significant in establishing an empirical benchmark for “should schedule” estimates: amount of Research Development Test and Evaluation (RDT&E dollars) at MS B start (in millions), the per cent of RDT&E funding at MS B start, whether a program is a modification and whether a program has a MS B start in 1985 or later. Although they explored and adopted significant variables to predict program schedule using pre-MS B data, they also considered a plethora of explanatory variables that were ultimately deemed statistically insignificant; we also consider these variables.

Brown *et al.* (2015) first identified the MS B start in 1985 or later as an explanatory variable. They demonstrate that programs with a MS B start date in 1985 or later have a statistically significant change in their expenditure profile. These programs tend to expend a greater percentage of their obligations by the program’s mid-point than the programs that start prior to 1985. Although not conclusive, Brown *et al.* (2015) hypothesized that the reason for this significant shift is owing to the President’s Blue Ribbon Commission on Defense (often referred to as the Packard Commission) and the acquisition reforms that occurred because of the recommendations of the commission.

Similar to Jimenez *et al.* (2016), Deitz *et al.* (2013) analyzed activities prior to MS B. They examined the importance of developing a robust analysis of alternatives prior to MS B and the effects that an analysis of this nature may have on program success. Their findings suggest that while only 10 per cent of a program’s life-cycle cost was invested prior to MS B, 70 per cent of a program’s lifecycle costs are committed by this milestone (Deitz *et al.*, 2013). This suggests to us that pre-MS B data may be very important to predicting program cost. However, this also limits data collection because pre-MS B reporting is not mandatory for all acquisition programs, and therefore, the cost and schedule data are unavailable in some instances. Jimenez *et al.* (2016) also experienced such a limitation.

Looking slightly further back in the literature, we find other pertinent studies that present possible explanatory variables to consider. Foreman (2007) researched methods to improve cost and schedule growth estimates by including longitudinal variables that account for changes that take place over time. His research built upon the database initially created by Sipple *et al.* (2004) and subsequently modified by Lucas (2004) and Genest and White (2005). Sipple *et al.* (2004) found the most important predictive variables of cost growth to be MS C to initial operational capability (IOC) duration and an indicator variable for a MS C slip.

The aforementioned researchers have identified numerous variables for investigation on whether they will be predictive of program cost. The complete list is in given the Appendix. This list also gives us our data inclusion and exclusion criteria. The initial data inclusion criteria include any program in the DoD (i.e. all service branches) which has reported program data using the selected acquisition reports (SARs). Additionally, they must be unclassified and reported within the Major Defense Acquisition Program (MDAP) and pre-Major Defense Acquisition Program (pre-MDAP) section of the Defense Acquisition Management Information Retrieval (DAMIR) database.

For a program to be considered in our study, it must satisfy three criteria. The first requirement is that the program SAR must contain an MS A date or funding at least one year prior to MS B – we interpret the pre-MS B funding as indicating the year in which MS A may have occurred. This requirement is because of the pre-MS B data being found predictive in the literature review. Unfortunately, this requirement also results in a great deal of programs being ineligible for inclusion because of a lack of reporting requirements prior to MS B. This is not unexpected considering a program is not official until meeting MS B.

We are able to include an additional 15 programs in our data set by making the following assumption when there is no MS A date provided: if there is funding in the funding profile at least one year prior to MS B, then MS A occurred in January of the year in which funding was first received. We did test this assumption to ensure these additional programs are not statistically different from the others prior to inclusion in the final data set.

The second exclusion criteria is that the program SAR must contain an MS B date and corresponding funding information. This again pertains to the necessity of containing pre-MS B data as a means to build a highly predictive model. Without the MS B date and funding information, we are unable to ascertain the duration of MS A or the funding spent up to MS B. Additionally, we are unable to calculate the projected funding needed to reach IOC or the projected duration of MS B to IOC.

The third exclusion criteria is that the program SAR must contain an IOC date that occurred prior to the last reported SAR which indicates that the program is complete up to IOC. This is important to our research, as it gives us a termination point to estimate and ensures we are not using projected values as actual values in our model. IOC is a very important date in a program, as it signifies the point in time when the program achieves an available capability in its minimum usefully deployable form.

As previously discussed, our data set starts with the 56 programs in the database built by Jimenez *et al.* (2016). We augment this database by analyzing defense program SARs from the DAMIR system. The program SARs contain program funding, schedule, and performance information relative to our research. Using our stated inclusion criteria, we add 187 programs to the initial 56. Then using the exclusion criteria, we remove 170 programs for a net change of 17; this results in a final program count of 73. Table I demonstrates inclusion and exclusion criteria used in this research. Table II lists the final 73 programs.

The data that we use for our analysis include both actual and projected values from the SARs. We use the latest available program’s SAR to record the actual cost from MS B to IOC as the response variable in the model. To develop a useful predictive tool for the acquisition community, we must only use projected cost and schedule data at MS B, as these are the only data the user of our regression model will have at their disposal.

To implement this limitation, we retrieve projected cost and schedule data from the SAR corresponding to the year in which MS B occurred or, if that SAR is unavailable, the earliest

Table I.
Program inclusion/
exclusion criteria and
counts

| Inclusion/exclusion criteria | Programs added | Program removed | Program count |
|------------------------------|----------------|-----------------|---------------|
| Jimenez’s starting database | 56 | | 56 |
| DAMIR query and addition | 187 | | 243 |
| Double count adjustment | | 29 | 214 |
| IOC occurs after last SAR | | 61 | 153 |
| Missing MS A or B | | 74 | 79 |
| Missing IOC | | 4 | 75 |
| Classified | | 2 | 73 |
| Total remaining | | | 73 |

| | | Modeling median will- cost estimates |
|-----|------------------------------|--|
| No. | Program | |
| 1 | A-10 | |
| 2 | AWACS | |
| 3 | C-17 | |
| 4 | F-22 | |
| 5 | AH-64 | |
| 6 | B-1B Computer upgrade | |
| 7 | C-5 RERP | |
| 8 | F-15 | |
| 9 | B-1B JDAM | |
| 10 | KC-135R | |
| 11 | FA-18 A/B | |
| 12 | AV-8B Harrier | |
| 13 | S-3A | |
| 14 | P-8 Poseidon | |
| 15 | V-22 Osprey | |
| 16 | E-2C Hawkeye | |
| 17 | F-35 JSF | |
| 18 | CH-47D Chinook | |
| 19 | E-8A JSTARS | |
| 20 | AGM-65A missile | |
| 21 | ALCM missile | |
| 22 | AMRAAM missile | |
| 23 | JASSM missile | |
| 24 | JDAM | |
| 25 | JPATS T-6A | |
| 26 | OTH-B | |
| 27 | LGM-118 peacekeeper | |
| 28 | GBU-39 SDB-I | |
| 29 | National aerospace system | |
| 30 | AGM-88 HARM | |
| 31 | AIM-9X Block 1 | |
| 32 | AN/BSY-1 | |
| 33 | Cobra Judy Replacement | |
| 34 | Harpoon missile | |
| 35 | NMT | |
| 36 | SH-60B | |
| 37 | UGM-96A Trident I missile | |
| 38 | SSN 774 (Virginia class sub) | |
| 39 | T-45TS | |
| 40 | UGM-109 Tomahawk | |
| 41 | SSBN 726 SUB | |
| 42 | AGM-114A Hellfire missile | |
| 43 | OH-58D Helicopter | |
| 44 | AAWS-M Javelin | |
| 45 | SSN 21 sub | |
| 46 | AWACS Blk 40-50 upgrade | |
| 47 | B-2 EHF Inc 1 | |
| 48 | C-5 AMP | |
| 49 | MQ-9 Reaper | |
| 50 | AH-64E remanufacture | |
| 51 | ATACMS-APAM | |
| 52 | CH-47F | |
| | (continued) | |

Table II.
List of programs
used in the research
database

Table II.

| No. | Program |
|-----|-------------------------|
| 53 | CSSCS (ATCCS) |
| 54 | Longbow Apache (AH-64D) |
| 55 | UH-60M Blackhawk |
| 56 | AESA |
| 57 | AGM-88E AARGM |
| 58 | CEC |
| 59 | E-2D AHE |
| 60 | JSOW |
| 61 | LCS |
| 62 | LHD-1 |
| 63 | MH-60R |
| 64 | MH-60S |
| 65 | Strategic sealift |
| 66 | Trident II |
| 67 | EA-6B ICAP III |
| 68 | JSIPS (CIGS) |
| 69 | NAS |
| 70 | AFATDS (ATCCS) |
| 71 | AEHF |
| 72 | EELV |
| 73 | WGS |

available SAR. This allows us to use projected values to predict a program’s cost from MS B to IOC, the same as if we were in a program office attempting to estimate the cost of our program independent of this research.

Methodology

To arrive at the presented model (explained in the next section), we use a mixed-direction stepwise approach to screen for the most predictive variables and then finalize the model using ordinary least squares (OLS). [Note: In statistics, stepwise is an analytical method of fitting regression models in which an automatic procedure chooses explanatory variables for addition or subtraction based upon a set criteria.] To eliminate the effects of inflation, we convert all funding variables to base year 2017 dollars (BY17) using the 2016 Office of the Secretary of Defense (OSD) inflation indices. For our regression model, the response variable is the natural log of the acquisition cost (defined as the RDT&E and Procurement costs) from MS B to IOC. We transform the response variable using a natural log function to mitigate against heteroskedasticity because of the large range of actual costs – without transforming the OLS residuals, we would have failed the assumption of constant variance at a level of significance of 0.05. To ascertain the actual cost estimate from the OLS model, we retransform the predicted output back to actual cost (in millions of BY 17 dollars) by calculating $e^{OLS\ Output}$. This transformed model results in a median estimate of will-cost, as this back-transformation equates to the median in the original space (Carroll and Rupert, 1981; Tisdell, 2006).

We use JMP® Pro 12 for our statistical analyses and adopt an initial overall experiment-wise Type I error of 0.1 owing to the exploratory nature of this study. To be consistent with this level of significance, we use a *p*-value threshold of 0.1 as the entry and exit criteria for the mixed direction stepwise regression model. Once the initial variables are identified by the stepwise procedure, we then use OLS to finalize the regression model. We now lower the overall Type I to be 0.05, and we require each predictor variable to be statistically significant according to the

Holm–Bonferroni method, which counteracts the problem of multiple comparisons (Holm, 1979). Prior to conducting the variable selection procedure, we randomly select 20 per cent, or 15, of the 73 programs and set these aside for utilization as a validation set. We use the remaining 58 programs for the stepwise and OLS regression analysis.

For our model to be considered viable, we must verify the standard OLS assumptions. To assess the assumptions of homoscedasticity and normality of model residuals, we conduct a Breusch–Pagan (B–P) and Shapiro–Wilk (S–W) test, respectively, at a level of significance of 0.05. To assess multicollinearity and possible influential data points, we examine the variance inflation factors (VIF) and evaluate Cook’s distance values, respectively.

After all the underlying model assumptions are assessed and passed, we test our resultant model against the validation pool (the 15 set aside programs) using descriptive and inferential measures. Regarding descriptive measures, we compute the absolute per cent error (APE) of the true cost between MS B and IOC and the predicted cost for each program. [Note: The true and predicted costs are evaluated in the natural log space.] Using these APE values, we then calculate the median and mean APEs (MdAPE and MAPE, respectively). We calculate these for both the validation and modeling programs and compare the values. We also investigate whether the untransformed predicted values truly reflect the median value or a baseline estimate for will-cost by investigating how the true program cost compare to the predicted program cost. After validating our selected model, we perform another mixed stepwise analysis using the entire data set of 73 programs to determine if we inadvertently left out a predictive variable.

Analysis

Using mixed stepwise regression on the modeling set of 58 programs, we develop a preliminary model. Table III highlights this model. The presented model has an R^2 , which represents the amount of variability in the data explained by the model, of 0.82. We calculate the APE values for this model which results in an MdAPE and MAPE of 0.050 (5.0 per cent) and 0.059 (5.9 per cent), respectively, for the model building set. For the validation set, we obtain an MdAPE and MAPE of 0.056 (5.6 per cent) and 0.079 (7.9 per cent), respectively. Although the validation set is slightly higher than the model building set, all of the absolute per cent errors are less than 10 per cent suggesting the model is performing well.

With respect to the inferential measures, Table III reveals that all VIF scores are below or close to 2, indicating little to no evidence for multicollinearity. The preliminary model also contains no Cook’s distance score above 0.50 (highest value is approximately 0.10). This suggests no overly influential data points affecting the p -values of our explanatory variables. Model residuals pass both assumptions of normality and homoscedasticity with p -values of 0.25 and

| Predictor variable | Estimate | P -value | Standardized estimate | Variance inflation factor |
|--------------------------------|----------|------------|-----------------------|---------------------------|
| Intercept | 5.731 | < 0.0001 | N/A | N/A |
| Projected MS B to IOC (months) | 0.0114 | 0.0033 | 0.199 | 1.13 |
| RDT&E \$ at MS B Start | 0.00029 | 0.0003 | 0.297 | 1.56 |
| Fixed wing | 0.620 | 0.0037 | 0.199 | 1.17 |
| Electronic system program | –0.732 | 0.0142 | –0.216 | 1.96 |
| ACAT I | 0.837 | 0.0346 | 0.160 | 1.47 |
| Large program | 0.747 | 0.0018 | 0.251 | 1.58 |
| Extra-large program | 1.205 | < 0.0001 | 0.397 | 2.16 |

Table III.
Preliminary OLS
model

0.92 for the S-W and B-P tests, respectively. Lastly, all explanatory variables are individually significant at the comparison-wise error rate under the Holm–Bonferroni criteria (Holm, 1979).

With the model being deemed internally valid, we combine all the data together to update model parameter values using OLS and lowering the overall Type I error rate to 0.05. Table IV shows the updated model. The stepwise approach failed to detect any additional predictor variables (at the overall familywise error rate of 0.05 level of significance), and the resultant model described in the next section is our final model. The resultant model has an R^2 of 0.83 with an MdAPE and MAPE of 0.057 (5.7 per cent) and 0.062 (6.2 per cent), respectively. This means that the presented model has a relative error of between 5.7 and 6.2 per cent of predicting the natural log of the program cost from MS B to IOC. After back-transforming to the original values of program cost from MS B to IOC, approximately 50.7 per cent of the 73 programs in our database had a true program cost exceeding the predicted cost while 49.3 per cent had less. Theoretically, this ratio should be 50 per cent by 50 per cent. The empirical percentages suggest our presented model is performing as expected.

To prevent model extrapolation, the ranges in which this model is useful for the two continuous variables must be consistent with the bounds of the programs used within our analysis. For projected duration from MS B to IOC the lower bound is 28 months while the upper bound is 129 months. For RDT&E funding (dollar million) at MS B Start (BY17), the lower bound is \$4.43m, while the upper bound is \$5,979.4m. Using this model outside of these ranges is inappropriate.

All of the statistically significant predictor variables are available to the cost estimator at the time the estimate is calculated (which is intended to be post-MS B). There is a limitation in the model in the sense that a “prior” cost estimate is required before engaging the presented model to help fine-tune the program cost estimate. However, we feel this limitation is minor given that the three binary variables that are cost-related, i.e. ACAT I, large and extra large, should be relatively certain as a program approached MS B:

- *(Projected) MS B to IOC Duration – continuous variable:* The parameter estimate of this variable is 0.0108 which is multiplied by the number of months the program estimates to spend from MS B to IOC. This duration does not necessarily correlate to the level of technology or technological maturity being used, but, rather, indicates the cost of time in DoD acquisition.
- *RDT&E funding (dollar million) at MS B start (BY17) – continuous variable:* The parameter estimate associated with this variable is 0.00026, which is multiplied by the actual, non-transformed RDT&E funding spent prior to program entrance into MS B. As the amount of funding spent at this point is additive to total program cost, we suggest that the amount of funding spent prior to MS B is indicative of the projected size and scope of the entire program. This variable could indicate a greater investment

Table IV.
Final ordinary least
squares model

| Predictor variable | Estimate | <i>p</i> -value | Standardized estimate | Variance inflation factor |
|--------------------------------|----------|-----------------|-----------------------|---------------------------|
| Intercept | 5.449 | < 0.0001 | N/A | N/A |
| Projected MS B to IOC (months) | 0.0108 | 0.0021 | 0.170 | 1.09 |
| RDT&E \$ at MS B Start | 0.00026 | 0.0007 | 0.220 | 1.50 |
| Fixed wing | 0.561 | 0.0039 | 0.165 | 1.19 |
| Electronic system program | −0.635 | 0.0061 | −0.191 | 1.77 |
| ACAT I | 1.151 | < 0.0001 | 0.251 | 1.38 |
| Large program | 0.758 | 0.0004 | 0.232 | 1.51 |
| Extra-large program | 1.461 | < 0.0001 | 0.439 | 2.12 |

in newer technology prior to MS B, which typically results in higher costs over the entire program life owing to integrating and further maturing this technology.

- *Fixed Wing – binary variable*: The parameter estimate associated with this variable is 0.561 and is multiplied by one for every aircraft (excluding helicopters) program estimate conducted. The positive parameter estimate indicates that aircraft programs sans helicopters appear to be more expensive in general in contrast to other DoD platform programs. We hypothesize this effect as an artifact of complexity associated with stealth, avionic, and engine capabilities of today's modern aircraft, regardless of branch of service.
- *Electronic system program – binary variable*: The parameter estimate associated with this variable is -0.635 and is multiplied by one for any program that is considered an electronic system program. The negative parameter estimate indicates these programs are statistically significantly cheaper to acquire than the other program types. [Bolten et al. \(2008\)](#) also concluded that electronic systems are historically cheaper.
- *ACAT I – binary variable*: The parameter estimate for this variable is 1.151 and is multiplied by a value of one for any program considered to meet ACAT I funding estimate requirements at the start of MS B. This variable being additive to program cost is logical owing to the nature of ACAT I programs and the dollar costs associated with these DoD acquisitions.
- *Large program – binary variable*: The parameter estimate for this variable is 0.758 and is multiplied by a one if the program being estimated projects to have a total program acquisition cost greater than \$7bn (BY17) (RDT&E and Procurement) from MS A to program conclusion but less than or equal to \$17.5bn (BY17). This value is estimated at MS B and was calculated using the 50 per cent interquartile from a histogram analyzing total projected program acquisition cost. The additive nature of this variable adjusts for large DoD acquisition programs.
- *Extra-large program – binary variable*: The parameter estimate for this variable is 1.461 and is multiplied by a value of one if the program acquisition cost from MS A to IOC is projected to be greater than \$17.5bn (BY17). This value is estimated at MS B and was calculated using the 75 per cent interquartile from a histogram analyzing total projected program acquisition cost. The additive nature of this variable adjusts for the largest DoD acquisition programs, such as the F-35 and F-22.

As an example of the model in action, suppose a program at MS B possessed the following characteristics: ACAT I, Fixed Wing, \$550m (BY17) of RDT&E funding at MS B start, and (Projected) MS B to IOC duration of 5 years (60 months). All of these values are within the observational window allowed by the model. Plugging those values into the model presented in [Table IV](#) and then backtransforming (via the natural exponent), results in a median will-cost estimate of approximately \$2.8bn (BY17) for MS B to IOC program acquisition costs. This value now serves as a benchmark to crosscheck should-cost estimates.

Discussion and conclusion

[Table V](#) presents the relative percentage contribution of each variable included in the final model. The smallest relative contribution is 9.9 per cent for fixed-wing aircraft, while the largest relative contribution is 26.3 per cent for extra-large programs. Besides these variables, there is low variation between the remaining predictor variables in the presented model. This suggests that the explanatory variables are relatively similar with respect to affecting the true program RDT&E and procurement costs.

Any statistical model has limitations. Principally, this model is based on data collected from SAR that sometimes contain incomplete information. Ultimately, the model is only as good as the data itself. The availability of pre-MS B data was a large constraint on the data building process and limited which programs could be included. Additionally, the search parameters used in DAMIR may have inadvertently removed useful programs from our study which might have influenced any number of other variables to be significant.

One significant limitation of the model is the high level of variability in the definition of IOC. Our model uses IOC as a termination point owing to the importance of this milestone in a program as well as the availability of the date. In the programs considered, the number of units considered for attaining IOC varies greatly. Achieving IOC is determined individually for each unique program based on an initial cadre of operators, maintainers and support equipment that can use and sustain the system in an operational environment. For example, satellite, submarine or ship programs may have IOC based on one single unit. Although in the case of missile programs, IOC could be in the hundreds of units. This drives a level of known variability within our model that could be better accounted for by using a more structured and universal definition for IOC; this could be a topic for future research.

Accurately predicting program cost is both an art and a science. Achieving accurate estimates during the early stages of a program’s lifecycle is an unenviable task, and one can be certain that the estimate will be wrong. However, deriving an estimate that is close to the final actual cost is crucial to improving the allocation of scarce resources. What our model provides is the empirical portion of the estimating process to ascertain the will-cost for a program. We provide this tool to the DoD acquisition community primarily as a method to check the assumptions and realism of their program office estimate. Being able to build a program cost estimate and turn to our statistically built and tested model for validation will be invaluable for the community because it will allow for an injection of increased realism into the cost estimating process. Realism in the will-cost median estimate is crucial to the success of should-cost analysis.

Drawing a difference between our research and prior research, the most notable difference is the model output. Our research and model focuses on building an empirically based estimate for program cost between MS B and IOC to serve as a realistic benchmark (the median value) for what programs will-cost. Program managers can then adopt “should cost” efficiencies to reduce cost further. We believe that modeling an output that will serve as an actual point estimate is valuable as a crosscheck tool for the user community. It gives the user a benchmark based on historical data against which the program measures its progress. The model also supports the “will-cost and should-cost” requirement levied in

Table V.
Predictor variables
and their relative
contribution to the
model

| Explanatory variable | Relative contribution (%) |
|--------------------------------|---------------------------|
| Projected MS B to IOC (months) | 10.2 |
| RDT&E \$ at MS B Start | 13.2 |
| Fixed wing | 9.9 |
| Electronic system program | 11.5 |
| ACAT I | 15.0 |
| Large program | 13.9 |
| Extra-large program | 26.3 |

2011 by providing an objective and defensible cost for what a program should actually cost based on what has been achieved in the past. Ultimately, a quality will-cost estimate provides a starting point for program managers to examine processes and find efficiencies that lead to reduced program costs.

References

- Birchler, D., Christle, G. and Groo, E. (2011), "Cost implications of design/build concurrency", *Defense Acquisition Research Journal*, Vol. 18 No. 3, pp. 237-246.
- Bolten, J.G., Leonard, R.S., Arena, M.V., Younossi, O. and Sollinger, J.M. (2008), *Sources of Weapon System Cost Growth: Analysis of 35 Major Defense Acquisition Programs (MG 670)*, RAND Corporation, Santa Monica, CA.
- Brown, G.E., White, E.D., Ritschel, J.D. and Seibel, M.J. (2015), "Time phasing aircraft R&D using the weibull and beta distributions", *Journal of Cost Analysis and Parametrics*, Vol. 8 No. 3, pp. 150-164, available at: <http://doi.org/10.1080/1941658X.2015.1096219>
- Carroll, R.J. and Rupert, D. (1981), "On prediction and the power transformation family", *Biometrika*, Vol. 68 No. 3, pp. 609-615, available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.536.2459&rep=rep1&type=pdf>
- Carter, A.B. and Mueller, J. (2011), "Should cost management: why? How?", *Defense At & L*, Vol. 40 No. 5, pp. 14-18, available at: www.dau.mil/pubscats/ATL%20Docs/Sep-Oct11/Carter_Mueller.pdf
- Deitz, D., Eveleigh, T.J., Holzer, T.H. and Sarkani, S. (2013), "Improving program success through systems engineering tools in the pre-milestone B acquisition phase", *Defense Acquisition Research Journal*, Vol. 20 No. 3, pp. 283-308.
- Foreman, J.D. (2007), "Predicting the effect of longitudinal variables on cost and schedule performance", Master's thesis, Defense Technical Information Center (DTIC) (ADA463520), Fort Belvoir, VA, available at: www.dtic.mil/get-tr-doc/pdf?AD=ADA463520
- Genest, D. and White, E. (2005), "Predicting RDT&E cost growth", *The Journal of Cost Analysis & Management*, Vol. 7 No. 1, pp. 1-12.
- Harmon, B.R. (2012), "The limits of competition in defense acquisition", paper presented at the Defense Acquisition University Research Symposium, Fort Belvoir, VA, available at: www.dau.mil/Research/symposiumdocs/Harmon-Nears substitutes paper.pdf
- Holm, S. (1979), "A simple sequentially rejective multiple test procedure", *Scandinavian Journal of Statistics*, Vol. 6 No. 2, pp. 65-70, available at: www.ime.usp.br/~abe/lista/pdf4R8xPVzCnX.pdf
- Jimenez, C.A., White, E.D., Brown, G.E., Ritschel, J.D., Lucas, B.M. and Seibel, M.J. (2016), "Using pre-milestone b data to predict schedule duration for defense acquisition programs", *Journal of Cost Analysis and Parametrics*, Vol. 9 No. 2, pp. 112-126, available at: <http://doi.org/10.1080/1941658X.2016.1201024>
- Lucas, B.M. (2004), "Creating cost growth models for the engineering and manufacturing development phase of acquisition using logistic and multiple regression", Master's thesis, Defense Technical Information Center (DTIC) (ADA422915), Fort Belvoir, VA, available at: www.dtic.mil/get-tr-doc/pdf?AD=ADA422915
- Sipple, V., White, E. and Greiner, M. (2004), "Using logistic and multiple regression to estimate engineering cost risk", *The Journal of Cost Analysis & Management*, Vol. 6 No. 1, pp. 67-79.
- Tisdell, J.E. (2006), "Small sample confidence intervals in log space back-transformed from normal space", Master's thesis, Defense Technical Information Center (DTIC) (ADA450276), Fort Belvoir, VA, Air Force Institute of Technology, available at: www.dtic.mil/get-tr-doc/pdf?AD=ADA450276

Appendix. Predictor variables investigated in this paper

MS A to MS B duration (months) – continuous variable

- According to the last SAR date, this variable indicates the total time it took in months for a program to complete from Milestone (MS) A to MS B. In this variable, we are only concerned with actual schedule duration data available to the cost estimator at the time of MS B/EMD start.

Quantity expected at MS B – continuous variable

- This variable indicates the estimate of total quantity of weapons systems that were expected to be produced at MS B at the time of the last SAR date.

RDT&E funding (dollar million) at MS B start (BY17) – continuous variable

- This variable is based on raw total RDT&E dollars (in millions) that were allocated to the program prior to MS B. The dollars were all standardized into the base year when the research began (BY17).

(Projected) per cent of RDT&E funding at MS B start (BY17) – continuous variable

- This variable is based on the percentage of available RDT&E dollars allocated to the program before, and up to the start of, MS B. While this variable is based on a percentage, the dollars that this percentage was derived from were all standardized into the base year when the research began (BY17).

(Projected) Total program acquisition cost (BY17) – continuous variable

- This variable is the total projected acquisition costs, from MS B to IOC, estimated at MS B or the earliest available program SAR. It serves to identify how large a program is projected to be in terms of cost.

Modification – binary variable

- This variable identifies programs whose existence serves as a modification to a pre-existing weapons system. If a weapons system is a modification, it does not necessarily mean it will not have pre-MS B data associated with it. Every program is different and, therefore, it cannot be assumed that a modification will automatically start at MS B.

Prototype – binary variable

- This variable identifies programs that create a prototype, or prototypes, of a weapons system before production of that weapons system begins. More than one type of prototype for a weapons system can be created in a given program.

Concurrency planned – binary variable

- This variable addresses planned concurrency in a given program prior to MS B. Concurrency is the proportion of RDT&E dollars that are authorized during the same years that Procurement appropriations are authorized. The planned level of concurrency forces managers to make decisions that can lead to [schedule] growth if either too much or too little concurrency is accepted for a given program (Birchler *et al.*, 2011, p. 246).

1985 or later for MS B start – binary variable

- This variable accounts for a time series trend of programs that started their MS B in 1985 or later. It is shown that programs which began development during 1985 or later (considered “contemporary”) expend a greater percentage of obligations by their schedule midpoint than the earlier pre-1985 programs. We attribute this difference to the President’s Blue Ribbon Commission on Defense (commonly called the Packard Commission) and the subsequent acquisition reforms.

Air Force – binary variable

- This variable identifies if the lead service on the program was the US. Air Force.

Navy – binary variable

- This variable identifies if the lead service on the program was the US. Navy.

Army – binary variable

- This variable identifies if the lead service on the program was the US. Army.

Marine Corps – binary variable

- This variable identifies if the lead service on the program was the US. Marine Corps.

Fixed wing – binary variable

- This variable identifies if the weapons system program is a fixed-wing aircraft program, regardless of service it is associated with. The criterion to qualify as a fixed-wing aircraft is for that weapons system to maintain flight via fixed wings versus rotary wing flight.

Fighter program – binary variable

- This variable identifies if the weapons system program is a fighter program, or close variation thereof, regardless of service it is associated with.

Bomber program – binary variable

- This variable identifies if the weapons system program is a bomber program, or close variation thereof, regardless of service it is associated with.

Helo program – binary variable

- This variable identifies if the weapons system program is a helicopter program, or close variation thereof, regardless of service it is associated with.

Cargo plane program – binary variable

- This variable identifies if the weapons system program is a cargo plane program, or close variation thereof, regardless of service it is associated with.

Tanker program – binary variable

- This variable identifies if the weapons system program is a tanker plane program, or close variation thereof, regardless of service it is associated with.

Electronic warfare program – binary variable

- This variable identifies if the weapons system program is an electronic warfare program, or close variation thereof, regardless of service it is associated with. An electronic warfare program, as not to be confused with an electronic system program, differs greatly in its main function(s). A description from Lockheed Martin makes the distinction that it involves the ability to use the electromagnetic spectrum – signals such as radio, infrared or radar – to sense, protect and communicate. At the same time, it can be used to deny adversaries the ability to either disrupt or use these signals (electronic warfare).

Trainer plane program – binary variable

- This variable identifies if the weapons system program is a trainer plane program, or close variation thereof, regardless of service it is associated with.

Missile program – binary variable

- This variable identifies if the weapons system program is a missile program, or close variation thereof, regardless of service it is associated with.

Electronic system program – binary variable

- This variable identifies if the weapons system program is an electronic system program, or close variation thereof, regardless of service it is associated with. This differs greatly from the previously described electronic warfare variable in that electronic systems programs are principally concerned with the electronic user interface of a system, avionics controls or other similar applications that primarily support the electronic usability of a system or system of systems.

Submarine program – binary variable

- This variable identifies if the weapons system program is a submarine program, or close variation thereof, regardless of service it is associated with.

Ship program – binary variable

- This variable identifies if the weapons system program is a surface ship program, or close variation thereof, regardless of service it is associated with.

Satellite program – binary variable

- This variable identifies if the weapons system program is a satellite program, or close variation thereof, regardless of service it is associated with.

ACAT I – binary variable

- This variable indicates if the program is an ACAT I program. This is significant, in that the ACAT I programs deal with a much larger dollar amount and are thus more susceptible to cost and schedule growth because of their large-scale nature and complexities.

(Projected) MS C to IOC duration (months) – continuous variable

- According to the earliest available SAR estimate, this variable indicates the total estimated time for a program to meet IOC from MS C. This variable has been found to be predictive of cost growth in the programs studied by [Foreman \(2007\)](#). With this variable, we are concerned with giving the cost estimator the ability to enter in the projected duration, in months, of the gap between MS C and IOC to predict program cost.

(Projected) MS C slip – binary variable

- This variable indicates whether the program projected date for meeting IOC extends past the initial estimate. [Foreman \(2007\)](#) has found that a slip in MS C is indicative of program cost growth in past research.

No MS A date – binary variable

- This variable identifies whether a program did not contain a MS A date in the schedule portion of the SAR, but it did include funding at least one year prior to MS B. This is used to identify these programs and test that they are not statistically different from the other programs and is not used in a predictive capacity.

Small program – binary variable

- This variable identifies whether a program's projected total acquisition costs (RDT&E and procurement) are below \$3bn. This value is determined from analyzing the histogram of the (projected) total program acquisition costs of the programs in our study and coincides closely with the 25 per cent value.

Medium program – binary variable

- This variable identifies whether a program's projected total acquisition costs (RDT&E and procurement) are above \$3bn but below \$7bn. This value is determined from analyzing the

histogram of the (projected) total program acquisition costs of the programs in our study and coincides closely with the 25-50 per cent range.

Large program – binary variable

- This variable identifies whether a program's projected total acquisition costs (RDT&E and procurement) are above \$7bn but below \$17.5bn. This value is determined from analyzing the histogram of the (projected) total program acquisition costs of the programs in our study and coincides closely with the 50-75 per cent range.

Extra-large program – binary variable

- This variable identifies whether a program's projected total acquisition costs (RDT&E and procurement) are above \$17.5bn. This value is determined from analyzing the histogram of the (projected) total program acquisition costs of the programs in our study and coincides with the 75 per cent value.

(Projected) per cent complete at MS B start – continuous variable

- This variable is motivated by the per cent of the RDT&E variable and serves to project the per cent that a program is complete, to IOC, when MS B occurs. It is calculated by dividing the projected duration from MS B to IOC by the sum of duration from MS A to IOC and projected duration from MS B to IOC. This serves to indicate where the program managers believe the program is in terms of schedule completeness. It could indicate program maturity level.

Corresponding author

Edward D. White can be contacted at: stat.associates@gmail.com