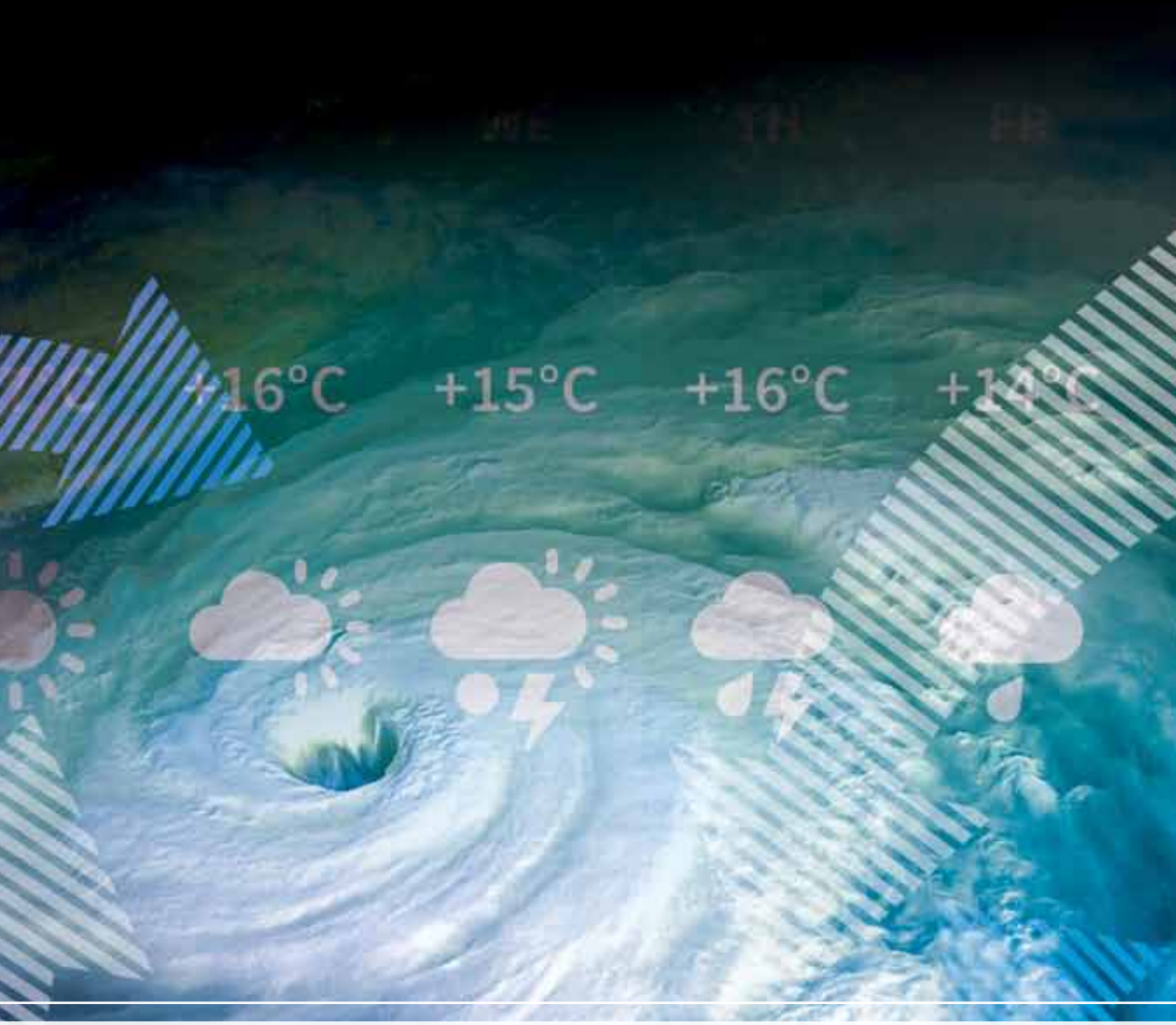
Taming the HURRICANE of ACQUISITION COST GROWTH —*Or At Least Predicting It*

Capt Allen J. DeNeve, USAF, Lt Col Erin T. Ryan, USAF, Lt Col Jonathan D. Ritschel, USAF, and Christine Schubert Kabban

Cost growth is a persistent adversary to efficient budgeting in the Department of Defense. Despite myriad studies to uncover causes of this cost growth, few of the proposed remedies have made a meaningful impact. A key reason may be that DoD cost estimates are formulated using the highly unrealistic assumption that a program's current baseline characteristics will not change in the future. Using a weather forecasting analogy, the authors demonstrate how a statistical approach may be used to account for these inevitable baseline

Image designed by Diane Fleischer



changes and identify related cost growth trends. These trends are then used to reduce the error in initial acquisition cost estimates by over one third for major defense acquisition programs, representing a more efficient allocation of $6 billion annually.

Keywords: *Capabilities Development for Rapid Transition, Joint Urgent Operational Need (JUON), Lethal Miniature Aerial Munition System, Operation Enduring Freedom, Operation Iraqi Freedom (OIF), Program of Record, Rapid Equipping Force*

The Not-So-Perfect Storm

Inaccurate cost estimates have long plagued Department of Defense (DoD) acquisition efforts. Despite the myriad acquisition reforms, and abundant detailed guidance on cost estimating best practices, accurately predicting the eventual cost of a weapon system remains difficult. A Government Accountability Office (GAO) study of all 96 active major defense acquisition programs (MDAP) in 2011 showed a total cost increase of over $74 billion in that year alone (GAO, 2012a)—an amount that would have paid for the 2013 defense sequestration cuts nearly twice over. The total MDAP portfolio cost continued to grow into 2013, despite a trend of reduction in the number of programs (GAO, 2014). A RAND study of completed major acquisition programs showed that the average cost estimate error measured from Milestone B is about 65 percent (Arena, Leonard, Murray, & Younossi, 2006a). This figure is an average of overestimates and underestimates; the absolute error is even higher. While researchers and practitioners may disagree on the efficacy of recent acquisition reforms upon improving cost estimates, clearly, there is ample room for improvement.

Perhaps the problem does not lie with the accuracy of the cost estimates, but with the fact that these estimates are accurately estimating the wrong thing. For example, when the RAND study corrected the cost data for changes in procurement quantity, the average cost errors dropped by over 20 percent (Arena et al., 2006a), and the GAO (2012a) study attributed nearly 40 percent of the $74 billion increase to quantity changes. If we expect accurate estimates of the final cost of acquisition programs, then we must take into account the uncertainty associated with program baselines upon which these estimates are based. We propose a method for correcting initial acquisition cost estimates using observed baseline deviations from similar past programs, thus reducing the average cost growth over these early estimates.

The Defense Acquisition University (DAU) defines cost growth as "the net change of an estimated or actual amount over a base figure previously established."[1] Many studies cite changes to the Acquisition Program Baseline (APB) as among the most significant sources of cost growth (Arena et al., 2006a; Drezner, Jarvaise, & Hess, 1993; GAO, 2012a). These studies often correct the cost estimates for these changes in an attempt to determine the programmatic causes for the cost overruns. In this way, researchers "maintain the integrity of the baseline" (Drezner et al., 1993, p. 11). These baseline-corrected analyses

> ***A more accurate prediction of the eventual cost of an acquisition program provides a better assessment of that program's affordability, thus better informing affordability decisions.***

are useful for driving acquisition reform, but they are less useful for informing resource allocation and affordability assessments, which are inherently more concerned with accurate prediction of actual program expenditures.

Will Cost, Should Cost, and Real Life

In a 2011 memorandum from the Assistant Secretary of the Air Force, Financial Management and Comptroller, and the Air Force Acquisition Executive (Department of the Air Force, 2011), the Air Force established the practice of generating two different cost estimates dubbed Will Cost and Should Cost. The Should Cost estimate is "based on realistic technical and schedule baselines and assumes success-oriented outcomes." In contrast, the Will Cost estimate is based on an independent estimate that "aims to provide sufficient resources to execute the program under normal conditions" (Department of the Air Force, 2011, p. 4). This notion that a program may cost something more than it should cost implicitly acknowledges that things don't always go as desired. Also, this concept sets the precedent that allowances may be made for difficulties through cost-estimating relationships that reference past development and production efforts as a benchmark.

In actuality, the Should Cost estimate does not incorporate enough realism. For example, common sources of cost growth, such as procurement quantity changes, are not included in the Should Cost estimate since this estimate is still based on the APB. This baseline specifies parameters such as procurement quantity, performance characteristics, program

duration, and so on. However, these baselines almost never remain constant (Drezner & Krop, 1997), leading inevitably to changes in program cost and crippling early estimating efforts.

A more accurate prediction of the eventual cost of an acquisition program provides a better assessment of that program's affordability, thus better informing affordability decisions. Therefore, the DoD needs a method for accurately estimating the final cost of an acquisition effort without relying on a fixed baseline. In this research, we have developed a novel method to correct early program cost estimates using high-level descriptive programmatic parameters. Advanced regression techniques establish a relationship between these parameters and the cost estimate error of past programs, and then use this relationship to predict estimate error in similar future programs. This method is dubbed "macro-stochastic" estimation (Ryan, Schubert Kabban, Jacques, & Ritschel, 2013, p. 3).

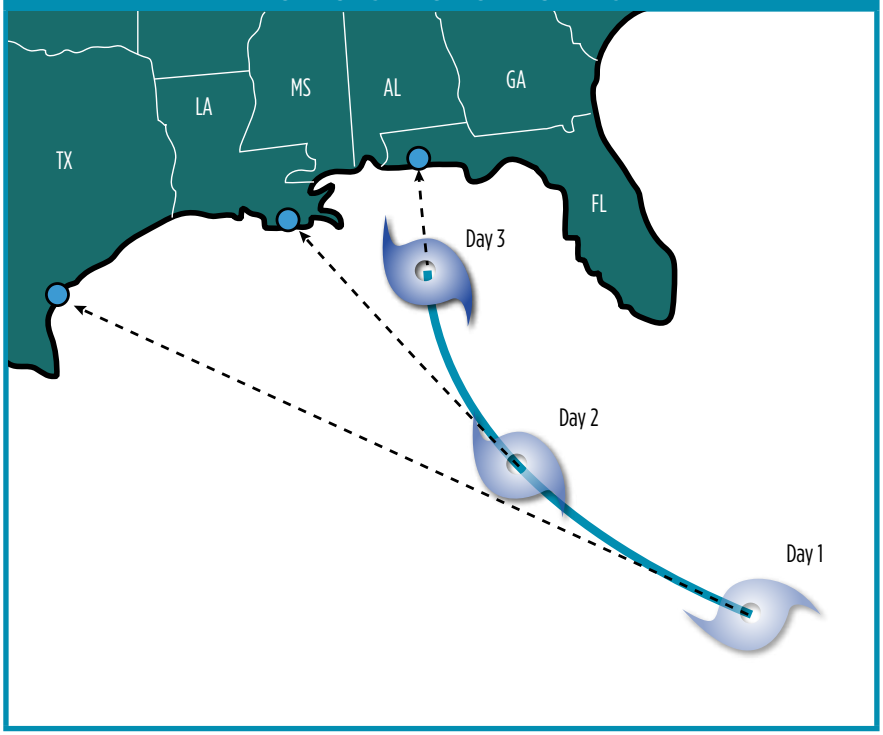
The National Oceanic and Atmospheric Administration (NOAA) uses a similar technique in the forecasting of hurricanes, a domain that has seen prediction accuracy triple in the last two decades (Silver, 2012). This fact is intriguing, because the challenges associated with predicting the path of a hurricane are remarkably similar to those of trying to predict and budget for the cost trajectory of a DoD program. In both cases, an extraordinary number of discrete, nonlinear elements all interact in exceedingly complex ways, serving to greatly complicate the task of predicting overall system behavior. And while the two phenomena both present similar estimating challenges, the modeling approaches and reporting conventions vary significantly.

We Know What a Bad Prediction Looks Like

For a moment, imagine that meteorologists forecast hurricanes in the same manner that the DoD budgets for acquisition programs. The local news channel reports that a hurricane has formed in the Caribbean. An expert team of meteorologists carefully examines the key characteristics of this newly formed hurricane, including its current location, size, speed, and heading. Based on this information, the meteorologists then officially announce their prediction for the hurricane: it will be a Category 2 hurricane that makes landfall at the intersection of Main Street and Third Avenue in Corpus Christi, Texas. The residents of Corpus Christi are notified of the threat. But, 24 hours later, the meteorologists follow this

same process, and provide an equally detailed—but vastly different—prediction. The Day 2 prediction is updated to take into account a new trajectory and larger size; now the storm is predicted to make landfall at the Northeast corner of the Walmart store in Cameron, Louisiana, as a Category 3 hurricane. The next day, this process repeats, predicting an even larger hurricane with a new landfall point in the parking lot of the Spinnaker Beach Club in Panama City, Florida. These volatile predictions are depicted in Figure 1.

FIGURE 1. AN ACCURATE (BUT LESS USEFUL) METHOD FOR FORECASTING HURRICANES



You might reasonably have many concerns about these estimates. For example, how likely is it that the hurricane will actually make landfall at these precise locations? You might wonder why each estimate only considers the current state of the hurricane as opposed to how it might change over time. And, of course, you might be highly skeptical of any set of estimates that varies so widely. But, this scenario does have some

unfortunate similarities with the DoD cost-estimating and budgeting processes. Although cost estimators carefully account for uncertainty in their cost estimates (based on a fixed APB), the official prediction is recorded into the budget as a point estimate. Their cost estimates typically include no consideration for a change in trajectory, and no indication of uncertainty in the eventual budget request. Just like in our fictitious forecasting scenario, we have an early prediction, but it is not a very good one since it is almost guaranteed to change. Updating the absurdly specific budget request at each milestone is not an adequate solution for addressing this change since substantial resources will have already been committed according to the original baseline. In fact, a common engineering adage presumes that 75 percent of the design cost is committed in the first 25 percent of the life cycle (Blanchard & Fabrycky, 2011).

Of course, this is not the way meteorologists forecast hurricanes. NOAA uses supercomputers running millions of advanced physics simulations to calculate the outcomes of minor changes in the weather's initial conditions, and these outcomes are combined to form a probabilistic prediction (e.g., "There is a 10 percent chance of rain today"). These simulations are supervised by experienced meteorologists, using their knowledge of past weather patterns to improve forecast accuracy by up to 25 percent over computer simulation alone (Silver, 2012). This marriage of cold calculations and "squishy" probabilistic judgments carries over to hurricane prediction; to predict the storm's path, NOAA uses this method of human-mediated simulation (Ferro, 2013).

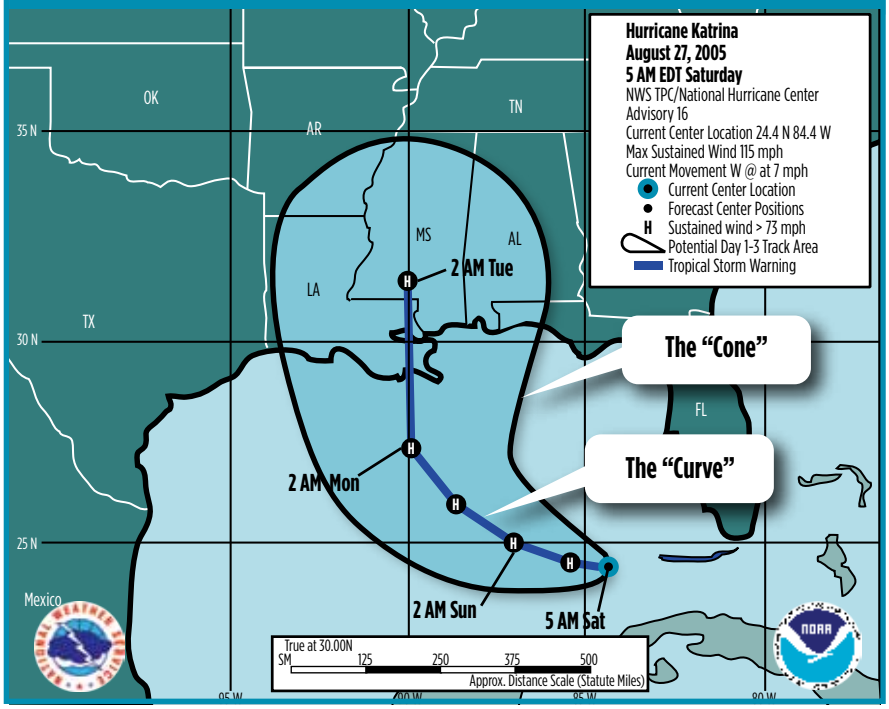
But for the prediction of hurricane strength, forecasters turn to what is essentially macro-stochastic estimation. They "compare basic information from the current storm, like location and time of year, to historic storm behavior," and use this information to predict the storm's strength (Ferro, 2013). In other words, top-level descriptive parameters are used to associate this storm with previous storms. The implicit assumption is that the current hurricane will perform similar to past hurricanes, as long as the right descriptive parameters are chosen. This combination of detailed simulation, coupled with statistical techniques (not to mention a healthy respect for uncertainty) produces the most useful estimate for informing evacuation decisions. That is, it results in a reasonably accurate prediction as early as possible.

However, embracing uncertainty is not synonymous with imprecision; for a prediction to be useful, it must not be overly vague. Most people are acquainted with the graphic that weather forecasters use to illustrate the expected path of hurricanes; an example is shown in Figure 2. This familiar visual form of prediction has two important elements:

1. The Cone: the region of uncertainty that shrinks as the storm approaches land and provides an idea of the confidence in the estimate.

2. The Curve: the change in trajectory that indicates the predicted path the storm will take.

FIGURE 2. NOAA HURRICANE TRACKER, SHOWING LANDFALL PREDICTION FOR HURRICANE KATRINA
Image Adapted from NOAA National Hurricane Center

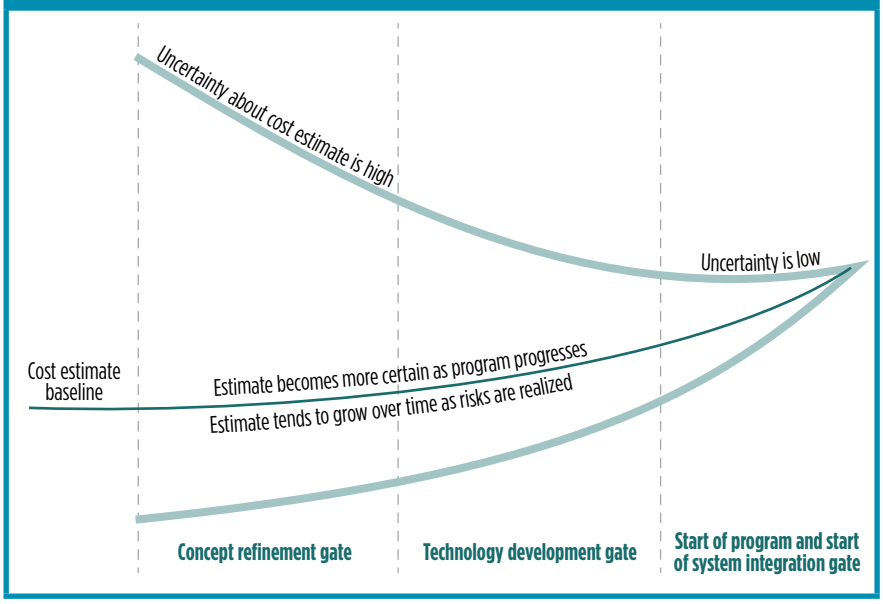


Note. Adapted from *Cost Estimating and Assessment Guide* (GAO-09-3SP), by Government Accountability Office, 2009, Washington, DC.

The Cone

The entire body of recent DoD cost-estimating guidance emphasizes the importance of risk analysis, sensitivity analysis, and the reporting of confidence in the program cost estimates (GAO, 2009; U.S. Air Force, 2007).[1] In fact, one might admire the similarity between NOAA's hurricane-tracking chart and a notional graphic from the *GAO Cost Estimating and Assessment Guide* (Figure 3) that illustrates the trajectory of a cost estimate baseline, with its accompanying cone of uncertainty (GAO, 2009). Unfortunately, the complex DoD process for turning an estimate into a budget does not possess a mechanism for incorporating uncertainty. Despite the best efforts of cost analysts to inform their customers of the confidence and possible risk in their calculations, these warnings are often interpreted as being too vague—a sentiment once expressed by an irate Harry S. Truman, who famously declared: "Give me a one-handed economist! All my economists say, 'on the one hand, on the other' " (Krugman, 2003). Incorporating uncertainty in budgeting activities requires a transformation in the way we think about resource planning. The first step in catalyzing such a revolution is likely to make provisions (or mandates) for reporting cost estimate

FIGURE 3. CONCEPTUALIZATION OF COST ESTIMATE TRAJECTORY AND CERTAINTY OVER TIME



uncertainty and confidence in acquisition status reports.[1] However, acquisition reform is beyond the scope of this study. Instead, we will focus primarily on "The Curve."

The Curve

It is not always reasonable to expect that the DoD can acquire a new weapon system for the Milestone B "sticker price." As one author recently noted, "Cost Discovery might be a better term for the process of updating estimates, because in retrospect it was clearly impossible to produce the stated capabilities for the original price" (Cancian, 2010, p. 396). It is rational to expect the rigors of research, development, and testing after Milestone B to uncover additional requirements that necessitate additional funding. But, if we are unable to completely avoid this "cost discovery," perhaps we should focus our efforts on predicting it. For example, consider the following questions:

- Is it true that an Air Force fighter aircraft program is likely to procure fewer aircraft than originally planned?
- Do Joint programs have significantly higher acquisition cost growth than non-Joint ones?
- Is the occurrence of a Nunn-McCurdy breach in a program a good indicator of future threshold breaches?

If we are able to hypothesize a relationship between these top-level program characteristics, then it is possible to examine past data to test if this relationship exists. Furthermore, if the relationship between these elements is, in fact, deemed statistically and practically significant, then we may apply this relationship to correct estimates in new programs. Macro-stochastic estimation is used to accomplish these goals.

Macro-Stochastic Estimation

To implement the macro-stochastic estimating technique described earlier, we first have to decide what high-level (macro) parameters are the most strongly associated with cost estimate errors. Next, we have to decide what constitutes a "similar program" so that we may apply the technique correctly on future data. In support of these pursuits, we have

created a database that tracks 75 distinct characteristics of MDAPs.[2] The Selected Acquisition Reports (SAR) for these programs are the source for our database.

Programs that have expended at least half of their planned funding are considered for entry in the database since these programs have sufficient data to measure trends in early program life. Also, only programs with a Milestone B date of 1987 or later are included. This cutoff date allows for a sufficient number of programs to estimate key characteristics and also maintains some continuity and relevance with current programs (Smirnoff & Hicks, 2007). This filtering process results in a sample of 937 SARs describing 70 programs from the Army, Navy, and Air Force. For each SAR, we compare the program's estimate of total acquisition cost against the actual cost specified in the program's final SAR. This ratio of estimated cost from a particular SAR to the final cost is defined as the Cost Growth Factor (CGF). For example, a program with a CGF of 1.3 indicates that the actual cost of the program was 30 percent higher than the original estimate. A program that perfectly estimated its final cost would have a CGF of 1.0.

A statistical technique known as mixed-model regression is applied to identify the parameters most strongly associated with changes in the final cost of a given program. This advanced statistical methodology is required due to the longitudinal nature of SAR analysis; that is, repeated measurements of the same program are expected to be correlated, violating a fundamental assumption of basic linear regression. Iteratively testing parameters in the dataset results in an efficient model of CGF containing the six parameters shown in Table 1.

It may seem like an oversight to omit an explanation of how each of these parameters affects CGF (that is, positively or negatively). In this case, the reason for this omission is related to the mixed-model methodology, and would surely have frustrated former president Truman, as the relationship varies depending on the program. Importantly, these six parameters are combined in different ways to create models tailored to specific groupings of programs, as described in the discussion that follows.

TABLE 1. SIGNIFICANT MODEL PARAMENTERS		
Parameter	**Description**	**Fixed/Variable**
Service Component	Identifies the executive military service (Army, Navy, or Air Force) that leads the acquisition program. Marine Corps programs are identified as belonging to the Navy.	Fixed
Development to Production Ratio	The ratio of the number of years a program spends in development to the number of years the program spends in production.	Variable
Count of Development APBs	This parameter tracks the number of times a new baseline is generated during the development phase.	Variable
Acquisition Cost	The total estimated program acquisition cost, as reported annually in the SAR.	Variable
Quantity Change	This parameter is tracked as a ratio of the procurement quantity planned in a given year to the original Milestone B procurement quantity.	Variable
Year Count	The sequential numbering of the program year, starting with Milestone B as year one. The presence of this parameter ensures the model is capable of predicting the estimate trends across time.	Fixed

Method

The mixed-model regression technique introduces flexibility that allows the analyst to generate different models for different groupings of programs. To return to our hurricane example, storms in the Caribbean might behave differently than those in the Atlantic. This difference may be taken into account by grouping the hurricane data into two bins, perhaps called Caribbean and Atlantic, and allowing the regression to generate separate estimates according to this partition. This feature is very powerful, since it can resolve patterns that might otherwise be

averaged out when the dataset is analyzed as a whole. More importantly, this feature allows us to bin acquisition programs into groups according to similarities in the behavior of their cost estimate error. When we wish to predict the CGF in a new program, we can apply the most appropriate model of estimate errors by determining the most suitable group for the new program.

The way programs are grouped is critical to the predictive power of the macro-stochastic technique. In theory, we could put all programs into the same group; but what we gain in broad model applicability, we sacrifice in accuracy. If the cost growth behavior for each of these programs was essentially the same, we wouldn't be so regularly thwarted when trying to produce a useful budget. Conversely, we could go with the opposite extreme and create a regression that examines each program individually by only assigning one program to each group. This grouping method results in a different model for each program and reduces nearly 99% of the error in program cost estimates! However, this accuracy is gained at the expense of utility. Future programs cannot be assigned to an existing group that is uniquely defined. The critical task, then, is to determine the most beneficial way to group the programs in order to balance accuracy with predictive capability.

Program Grouping

In this study, programs are grouped according to the categorical variables that are most strongly correlated with the CGF. These variables are simply characteristics of the program that are known in the first year, and reported in the first SAR. For example, final cost growth tends to be higher for new-start programs than programs that are essentially modifications or variants of existing weapon systems. Therefore, identification of program iteration is used to distinguish program groupings. The implicit assumption with this approach is that programs with similar overall cost growth will also exhibit similar cost growth patterns. The variables selected to bin programs are defined below.

1. Program Type. Based on the program description in the SAR, each program is placed into one of seven categories: Aviation, Electronic, Ground Vehicle, Maritime, Munition, Space, and Space Launch. These categories are consistent with previous program type categorizations (Arena et al., 2006a; Drezner et al., 1993).

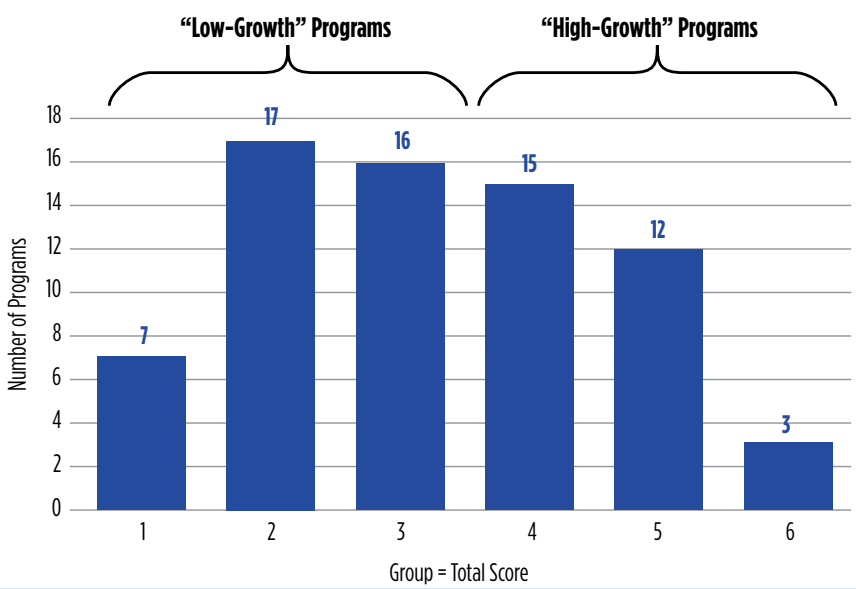
2. Iteration. This variable states whether a program is new, a lettered-variant on an existing program (e.g., the F-16 C/D), or a modification to an existing program (e.g., the C-5 Avionics Modernization Program).

3. Number of Years Funded. This variable describes the number of years the program is expected to be funded. This variable may change due to funding volatility.

4. Joint. This binary variable indicates whether a program is Joint between two or more Services.

Program groups are created by dividing each of the variables into levels, ensuring sufficient sample size within each level. A program is assessed a CGF "score" based on the applicable level for each of the four variables. The program group is the sum of the CGF scores across the four variables. Each program is scored in this manner, and the total scores from each program form the six program groups shown in Figure 4.[3]

FIGURE 4. PROGRAM GROUPS RESULTING FROM CGF SCORES



Validation and Results

The mixed-model regression uses the program groups in Figure 4 to fit different models using the significant CGF predictors shown previously in Table 1. However, due to relatively few programs in certain groups, validating the model is necessary without omitting too many of our samples for this purpose. Consequently, we validate the model using a technique that omits program data in a round-robin fashion, predicting the CGF of the omitted program and then replacing the data to make the prediction for the next omitted program. This validation is a type of Leave One Out Cross-Validation tailored to multilevel or mixed models (Ryan et al., 2013). It results in the aggregation of 70 separate analyses (one for each program) into a single set of results that reflects the expected predictive power of the macro-stochastic model. The validated model produces a set of predicted CGFs for every program estimate throughout the life of every program in our sample. If this version of the model is deemed reasonably powerful, then the original fitted model is considered validated and is the final model reported for inference.

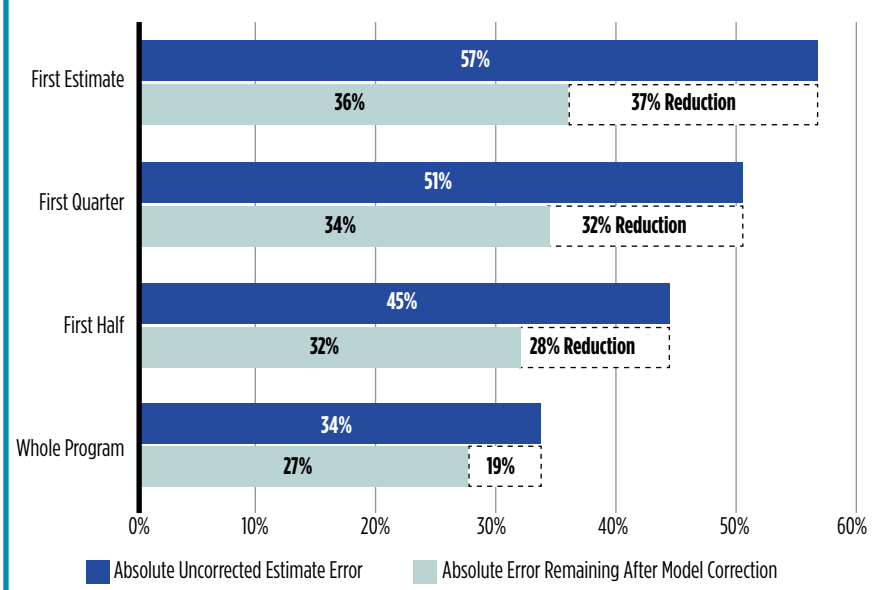
Using the validated results, the predicted CGF for any SAR that meets the established completion criteria may be used to correct the cost estimate in that SAR, but some of these corrections will be more useful than others. Since the SAR estimates get progressively better over time, there is equivalently less CGF error for the model to correct, thus reducing the average predictive performance of the model as a program matures. Consequently, the macro-stochastic technique is most useful when applied to correct the earliest cost estimates in a program. In fact, for each additional percentage of program expenditure, the model loses approximately three-quarters of a percent of its predictive power.

The 70 programs in our dataset displayed a mean CGF of 1.44, measured from the initial SAR estimate. This means that the programs underestimated their eventual cost by 44%, on average. However, this is an average of underestimates and overestimates. For the purposes of resource allocation, under and overestimation of budgetary requirements may both be considered detrimental because dollars allocated to one program cannot be easily transferred to another. Since the model seeks to minimize

cost estimate errors regardless of direction, the absolute estimate error is a more appropriate measure. Our sample showed a mean absolute error of 57%.

In contrast, after applying the macro-stochastic technique, the model-corrected CGF for these initial estimates averaged 0.93—slightly over-estimating, but closer to the ideal 1.0 CGF. As shown in Figure 5, the average absolute error for model-corrected estimates was 27%, representing a 19% reduction in the average absolute cost estimate error, across all programs. However, model performance is best in early program life; the average error reduction in the first estimate is 37%. Also, since the six program groups are assigned by assessing the severity of their cost growth, we expect that the most significant improvement will be seen when the model is applied to the "high-growth" programs. When the algorithm is applied to the first estimate of programs in CGF categories four through six, 90% of these estimates are improved, with an average error reduction of 45%.

FIGURE 5. SUMMARY OF VALIDATED MODEL PERFORMANCE ACROSS ALL 70 PROGRAMS



Reporting model performance as a percent improvement is useful because it normalizes programs of disparate cost. However, since our research focuses on real dollars, it is important to convert the percent error reduction into a dollar amount to demonstrate model efficacy. The absolute percent error for each program is multiplied by its final cost and converted to base year 2013 dollars in order to establish the total dollar amount reallocated by the validated model. The aforementioned 19% reduction in error equates to $119.5 billion, in base year 2013 dollars. If the total cost of these programs is scaled to equal that of the current DoD MDAP portfolio (DoD, 2013), then this macro-stochastic model could potentially allocate $6.24 billion more efficiently every year, if consistently applied to the first estimate of new MDAPs.

What This Technique Is Not

These results clearly illustrate the utility of the macro-stochastic cost-estimating approach. But, as is often the case with statistical tools, it is perhaps equally important to manage expectations by explaining a few of the applications for which this technique is ill-suited.

1. Adjusting cost estimates at the program office level. The efficacy of the model deteriorates rapidly and, even when applied to the first estimate of every program, only about 72% of program estimates are improved. This notion that estimates are only improved on average can be a significant source of doubt when it suggests that a program's rigorously developed estimate might be 44% too low. However, the average cost of programs is sufficient for informing better affordability decisions when considering a portfolio of assets.

2. Placing blame and driving acquisition reform. Macro-stochastic estimation eschews the typical cause-and-effect relationship that so many other acquisition studies seek to uncover. Rather, the model draws its power from the correlation between seemingly unrelated things. For example, it would be incorrect to say that the Service Component causes cost growth; it is simply an observed correlation. This lack of causality makes this model ill-suited for suggesting changes to the acquisition process.

3. Placing bounds on a traditional cost estimate. The full text of this study (DeNeve, 2014) explains the prediction intervals that surround the estimates of CGF. However, these alone do not constitute the "cone of uncertainty" discussed earlier in this article. With changes to the APB, the distribution around the predicted CGF and the cost estimate will change. Both of these distributions must be taken into account when placing bounds on the model-corrected final cost estimate. This is a subject for future work.

Conclusions

The existing paradigm for reporting acquisition cost based on a fixed APB results in unrealistic budgets and chronically inefficient resource allocation. In the current environment of fiscal restraint, embracing uncertainty can help provide a more realistic view of a program's true affordability. Acknowledging the likelihood of changes to a program's baseline grants the freedom to leverage past data and predict trends in cost-estimate performance. While not suitable as a low-level cost estimating tool, this study demonstrates such a method to reduce cost-estimate error in the earliest estimates of major defense programs, helping to stabilize long-term, portfolio-level budgets. As demonstrated by Figure 5, our model achieves the most significant error reduction early in program life, when accurate estimates are crucial for resource allocation and affordability decisions. In fact, nearly half of the estimate error is reduced when the model is applied early to the most growth-prone acquisition programs. As with hurricane forecasting, the optimal approach for acquisition cost estimation is likely a combination of techniques that focuses on providing the most useful estimate, even if this means embracing the uncertain nature of defense acquisition.

References

Arena, M. V., Leonard, R. S., Murray, S. E., & Younossi, O. (2006a). *Historical cost growth of completed weapon system programs*. Santa Monica, CA: RAND Corporation.

Blanchard, B. S., & Fabrycky, W. J. (2011). *Systems engineering and analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Cancian, M. F. (2010). Cost growth: Perception and reality. *Defense Acquisition Review Journal*, *17*(3), 389–403.

DeNeve, A. (2014). *A macro-stochastic approach to improved cost estimation for defense acquisition programs* (Master's thesis). AFIT-ENV-14-M-20. School of Engineering and Management, Air Force Institute of Technology, Wright-Patterson AFB, OH.

Department of the Air Force. (2011). *Implementation of will-cost and should-cost management* [Memorandum]. Washington, DC: Offices of the Assistant Secretary of the Air Force (Financial Management & Comptroller) and Air Force Acquisition Executive.

Department of Defense. (2013). *National defense budget estimates for FY 2014*. Washington, DC: Office of the Under Secretary of Defense (Comptroller).

Drezner, J., Jarvaise, J., & Hess, R. (1993). *An analysis of weapon system cost growth* (Report No. MR-291-AF). Santa Monica, CA: RAND Corporation.

Drezner, J., & Krop, R. (1997). *The use of baselining in acquisition program management*. Santa Monica, CA: RAND Corporation.

Ferro, S. (2013). Why it's so hard to predict hurricanes. *Popular Science*. Retrieved from http://www.realclearscience.com/2013/08/30/why_it039s_so_hard_to_predict_hurricanes_254806.html

Government Accountability Office. (2009). *Cost estimating and assessment guide* (Report No. GAO-09-3SP). Washington, DC: Author.

Government Accountability Office. (2012a). *Defense acquisitions: Assessments of selected weapon programs* (Report No. GAO-12-400SP). Washington, DC: Author.

Government Accountability Office. (2012b). *Defense logistics: Improvements needed to enhance oversight of estimated long-term costs for operating and supporting major weapon systems* (Report No. GAO-12-340). Washington, DC: Author.

Government Accountability Office. (2014). *Defense acquisitions: Assessments of selected weapon programs* (Report No. GAO-14-340SP). Washington, DC: Author.

Krugman, P. (2003). *The one-handed economist*. London: Economist Newspaper Ltd.

Ryan, E., Schubert Kabban, C., Jacques, D., & Ritschel, J. (2013). A macro-stochastic model for improving the accuracy of Department of Defense life cycle cost estimates. *Journal of Cost Analysis and Parametrics*, *6*(1), 43–74.

Silver, N. (2012). *The signal and the noise*. New York: Penguin.

Smirnoff, J., & Hicks, M. (2008). The impact of economic factors and acquisition reforms on the cost of defense weapon systems. *Review of Financial Economics*, *17*(1), 3–13.

U.S. Air Force. (2007). *U.S. Air Force cost risk and uncertainty analysis handbook*. Bedford, MA: Telecote Research.
Weapon Systems Acquisition Reform Act of 2009, 10 U.S.C., Pub. L. 111-23 (2009).

Endnotes

[1]The *Defense Acquisition Guidebook* dictates that MDAPs "must state the confidence level used in establishing a cost estimate…in the next Selected Acquisition Report prepared in accordance with 10 U.S.C. § 2423" (DAU, n.d., Chap 3, §3.4.1). The referenced section of U.S. Code contains no such requirement, and few SARs currently report confidence in their estimates.

[2]MDAPs are the largest programs in the DoD, defined by having more than $509 million for Research, Development, Test & Evaluation, or more than $3 billion for procurement in Base Year 2010 dollars (Weapon Systems Acquisition Reform Act, 2009). In fiscal year 2014, MDAPs constituted 40 percent of the acquisition funding for the DoD (DoD, 2013) and since 1969, they have been required to submit a standardized annual report of their status, called the Selected Acquisition Report (GAO, 2012b).

[3]This scoring methodology is explained in far greater detail in the full text of the study (DeNeve, 2014).

Author Biographies

Capt Allen J. DeNeve, USAF, is an engineer at the Air Force Life Cycle Management Center. Capt DeNeve's professional experience to date includes systems engineering and acquisition program management. His principal research interests are systems engineering, modeling and simulation, and operations research. Capt DeNeve holds an MS in Systems Engineering from the Air Force Institute of Technology, and a BS in Aeronautical Engineering from the U.S. Air Force Academy.

(E-mail address: allen.deneve.1@us.af.mil)

Lt Col Erin T. Ryan, USAF, is an assistant professor in the Department of Systems Engineering and Management at the Air Force Institute of Technology (AFIT). Lt Col Ryan's principal experience has been in the intelligence and space communities. His research interests include stochastic modeling, decision analysis, life cycle cost estimating, and space architectures. Lt Col Ryan holds a PhD in Systems Engineering from AFIT, and degrees in Electrical Engineering and National Security Studies from the University of Washington and New Mexico State University, respectively.

(E-mail address: erin.ryan@afit.edu)

Lt Col Jonathan D. Ritschel, USAF, is an assistant professor and director, Cost Analysis Program, in the Department of Systems Engineering and Management at AFIT. His research interests include public choice, the effects of acquisition reforms on cost growth in DoD weapon systems, research and development cost estimation, and economic institutional analysis. Lt Col Ritschel holds a PhD in Economics from George Mason University, a BBA in Accountancy from the University of Notre Dame, and an MS in Cost Analysis from AFIT.

(E-mail address: jonathan.ritschel@us.af.mil)

Dr. Christine Schubert Kabban is an assistant professor of statistics in the Department of Mathematics and Statistics at AFIT. She rejoined AFIT after 5 years in the Department of Biostatistics at Virginia Commonwealth University. Her research interests include classification and detection methods, diagnostic testing, information fusion, and multilevel modeling. Dr. Schubert Kabban holds a PhD in Applied Mathematics from AFIT, an MBA in Finance, and an MS in Applied Statistics from Wright State University.

(E-mail address: christine.schubert@afit.edu)