SYM-AM-19-070

PROCEEDINGS
OF THE
SIXTEENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

THURSDAY SESSIONS
VOLUME II

Acquisition Research:
Creating Synergy for Informed Change

May 8–9, 2019

Published: April 30, 2019

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

Quantifying the Year-by-Year Cost Variance of Major Defense Programs

David M. Tate—joined the research staff of the Institute for Defense Analyses' (IDA) Cost Analysis and Research Division in 2000. Prior to that, he was an Assistant Professor of Industrial Engineering at the University of Pittsburgh, and the Senior Operations Research Analyst (Telecom) for Decision-Science Applications, Inc. At IDA, he has worked on a wide variety of resource analysis and quantitative modeling projects related to national security. These include an independent cost estimate of Future Combat Systems development costs, investigation of apparent inequities in Veterans' Disability Benefit adjudications, and modeling and optimization of resource-constrained acquisition portfolios. Dr. Tate holds bachelor's degrees in philosophy and mathematical sciences from the Johns Hopkins University, and MS and PhD degrees in operations research from Cornell University. [dtate@ida.org]

Michael R. Guggisberg—is a Research Staff Member in the Strategy, Forces, and Resources Division at the Institute for Defense Analyses. He earned his PhD in economics at the University of California, Irvine in 2017. Prior to that, he was a squad leader in the United States Marine Corps Reserve, with tours in Iraq and Afghanistan. [mguggisb@ida.org]

Abstract

To a first approximation, acquisition programs never spend what they originally said they would spend when they began. In fact, the uncertainty in initial funding profile estimates is much larger than is generally understood; the possibility of program cancellations, restructurings, truncations, and block upgrades are often not accounted for. Worse yet, all of this uncertainty arises in a context in which programs must fit within annual budgets—it is not enough to only spend as much as you said you would; you must also spend it when you said you would, or problems ensue.

In 2018, we presented a methodology that uses historical program outcomes to characterize the year-by-year development and procurement cost risk associated with a major acquisition program. That work used functional regression to characterize changes in development profiles, modeled as Weibull curves. This paper improves and extends that work, using a novel application of Functional Principal Component Analysis (FPCA) to characterize the distributions of future RDT&E and Procurement profiles of both new and continuing acquisition programs.

Introduction—The Research Program

Recap of Prior Work

To a first approximation, acquisition programs never spend what they said they would when they began. In fact, the error bars around an initial cost estimate are much larger than is generally understood once program cancellations, restructurings, truncations, and block upgrades have been accounted for. Worse yet, all of this uncertainty arises in a context where programs must fit within annual budgets—it is not enough to only spend as much as you said you would; you must also spend it when you said you would, or problems ensue.

We have developed a methodology to characterize the year-by-year budget risk associated with a major acquisition program. This methodology can be applied to both development costs (Research, Development, Test, and Evaluation, or RDT&E) and procurement costs, and can be extended to understand the aggregate affordability risk of portfolios of programs. The method allows resource managers to estimate annual budget risk levels, required contingency amounts to achieve a specified probability of staying within a given budget, and a host of other relevant risk metrics for programs. It also allows policy makers to predict the impact on program affordability of proposed changes in how contingency funds are managed.

Many researchers have studied cost growth in major defense programs. The vast majority of this work has looked at either the ratio of eventual total cost to the originally estimated total cost, or the increase in some unit cost measure. Neither of these approaches addresses the problem that funds are authorized year-by-year, and that the affordability of a program or portfolio requires having enough obligation authority in each year to do the work needed over the next few years.

In Tate, Coonce, and Guggisberg (2018), we introduced an analytical approach for quantifying how likely a given set of programs is to fit within a projected budget over a planning horizon. This paper improves and extends that previous work.

To recap the approach: using historical Selected Acquisition Report (SAR) data, we look at how the profile of annual funding changed from initial estimates to actual authorized amounts, looking only at programs that are no longer spending. We do this separately for RDT&E costs and procurement costs. Our approach is agnostic about causes of these changes—the possibility that a program might be cancelled, or that the buyer might decide to triple the quantity or modify the design, is treated as part of the uncertainty to be accounted for in forecasting future budget demands. Posterior estimates of the distribution of possible cost profile outcomes are generated as a function of initial budget estimates, attributes of the program (e.g., that it is a joint program, or an aircraft program), and environmental conditions (e.g., that overall defense budgets are relatively tight at the moment).

Reminder: Desirable Outputs of a Model

Given a planned program (or set of programs—we'll get to that later) and a budget, resource managers would very much like to answer questions such as the following:

- What is the distribution of funding the program will receive in year t = 1, 2, ?
- What is the probability that the program will receive more funding in year t than is currently budgeted, for t = 1, 2, ?
- How many total contingency dollars would be enough to achieve a given probability that the current budget plus the contingency is enough to fund the program over the Future Years Defense Plan (FYDP)?
- What is the probability that the program will receive at least $X less than planned over the FYDP, for various values of X?

The goal of our research is to develop empirical models, based on historical program attributes, environments, and outcomes, that will allow us to answer questions like these. To do that, we need a few specific tools:

- A way to describe funding profiles mathematically;
- A list of program attributes and environmental factors that help predict program outcomes;

- A statistical model to estimate the probability distribution of final funding profile shapes, given the initial or midlife funding profile, environmental factors, and other program attributes;
- A mathematical characterization of how well the shape tends to fit actual data; and
- Historical data on program initial plans, midlife plans/outcomes and final outcomes.

Tate et al. (2018) illustrated this approach using Weibull curves to model RDT&E cost profiles of major defense acquisition programs (MDAPs). That work used *functional regression*, in which the shape of the realized cost profiles is assumed to have a particular functional form. Specifically, we assumed a Weibull distribution for the (scaled) initial and final profiles. That approach proved to be unsatisfactory in a couple of ways.

For one thing, many historical programs have realized RDT&E spending profiles that do not look like a single Weibull profile. For example, the Advanced Medium-Range Air-to-Air Missile (AMRAAM) and DDG-51 Destroyer programs each consists of a sequence of block upgrades (or new developments) of the product. They behave, in essence, like multiple sequential acquisition programs under a single funding line.

Alternatively, some RDT&E programs function more like services contracts than like product development contracts, consisting of a level of effort to improve capabilities over time, rather than one or more development and production projects with a discrete beginning and end. Ballistic Missile Defense System (BMDS) is perhaps the best exemplar of this approach, but there are others. The Evolved Expendable Launch Vehicle program was originally designed as a program to procure a set number of launch vehicles for satellites. Now called the National Security Space Launch program, it represents ongoing modernization and improvement of space launch capabilities. Figure 1 shows annual RDT&E funding for BMDS.

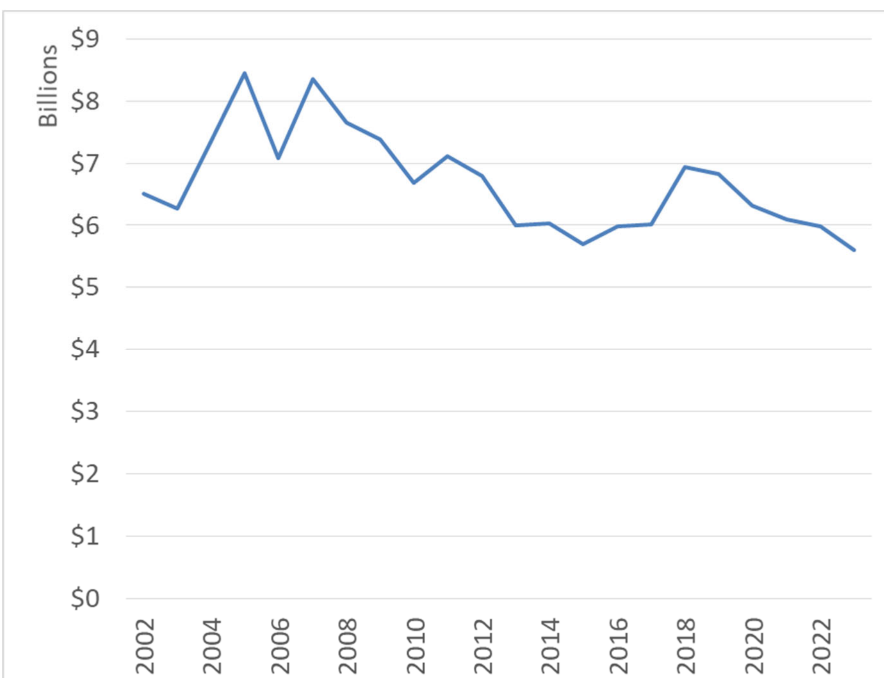

Figure 1. Annual RDT&E Funding for BMDS

Improved Modeling Approach

In the previous section, we noted that parametric functional families (and Weibull curves in particular) lack the flexibility to capture the variety of shapes shown by historical funding profiles. Our examination of historical spending patterns suggests that this is true not only of RDT&E profiles, but also of procurement profiles. As a result, we have adopted a nonparametric approach to characterizing profile shapes.

Instead of treating the year-by-year outcomes as having some complicated joint distribution, we will instead use nonparametric techniques from functional data analysis to treat the individual year-by-year outcomes as having been generated by some (noisy) underlying set of basic profile shapes, and then think about probability distributions over the parameters of those generating functions. Brown et al. (2015) provide a good summary of past approaches. Our first attempts (reported in Tate et al., 2018) attempted to fit the cost profiles to a pre-specified parametric functional family such as Weibull curves. Our revised approach uses a more flexible methodology based on Functional Principal Component Analysis (FPCA), described next.

Functional Principal Component Analysis

Define the set of programs to be $\{1,2,\ldots,I\}$. Let $C_{ij}(t)$ represent the planned spending for program i in fiscal year t as estimated in fiscal year j. Elements of $C_{ij}(t)$ reflect predictions if $t > j$ or actual spending for year t if $t \leq j$. This definition of a program can capture all stages of a program's lifecycle. A program is defined as initial if $C_{ij}(t) = 0$ for all $t < j$. A program is defined as completed (possibly cancelled) if $C_{ij}(t) = 0$ for all $t \geq j$. All other programs are considered "midlife"—their cost profiles are partly realized, but not yet completed.

A set of functional observations is notoriously difficult to summarize, since they are elements of an infinite dimensional space. One tool for summarizing such collections of functions is FPCA. FPCA is the infinite dimensional generalization of Principal Component Analysis, which is a methodology that provides an orthonormal basis to represent vectors in a finite dimensional Euclidean space. Its principal use in statistics is to find transformations of the predictive variables that are approximately independent in their effects on the outcomes of interest.

The FPCA process identifies a mean function $\mu(t)$ and a set of K eigenfunctions, $\xi_k(t)$, for $k \in \{1,2,\ldots,K\}$, that represent recurring patterns of deviation from the mean function. The eigenfunctions form an orthonormal basis in the L^2 Hilbert space (Yao, Müller, & Wang, 2009). The FPCA basis explains more variation than any alternative basis expansion when using a fixed K number of eigenfunctions.

Given an observed historical cost profile $C_{ij}(t)$, FPCA represents the profile as a weighted sum of the eigenfunctions, plus the mean function:

$$\log\left(C_{ij}(t)\right) = \mu(t) + \sum_{k=1}^{K} \omega_{ijk}\xi_k(t) + \epsilon_{ij}(t).$$

That is, ω_{ijk} (usually called "FPCA scores") function as weights on the eigenfunctions for generating log cost profiles in the new basis. The mean function and eigenfunctions are common for all programs in all stages of their life. The kth FPCA score is specific to program i in fiscal year j. The discrepancy from using a fixed finite K number of eigenfunctions is represented by $\epsilon_{ij}(t)$.

We apply FPCA separately to RDT&E and procurement cost profiles, for several reasons:

- RDT&E spending profiles and procurement spending profiles tend to have different shapes;
- RDT&E profiles and procurement profiles are offset in time, with procurement spending beginning later; and
- RDT&E and procurement fall under different "colors of money," and must therefore be separately evaluated against their respective budgets.

We use $K = 3$ for RDT&E and $K = 2$ for procurement. The value K was chosen such that the cumulative fraction of explained variation was over 90%. Applying FPCA generates mean profiles and principal eigenfunctions for both RDT&E and procurement. The shape fits are done using profiles that have been scaled in duration such that they begin at time 0 and end at time 1. The mean and principal eigenfunctions are shown in Figure 2.. The mean function is in the upper two subplots and the first K principal eigenfunctions are in the bottom two subplots. Note that while the mean RDT&E profile does have a roughly Weibull shape, the FPCA method can also account for more complex shapes using different weights on the various eigenfunctions. This is an improvement over the previous method, which would force a Weibull shape even where not appropriate.

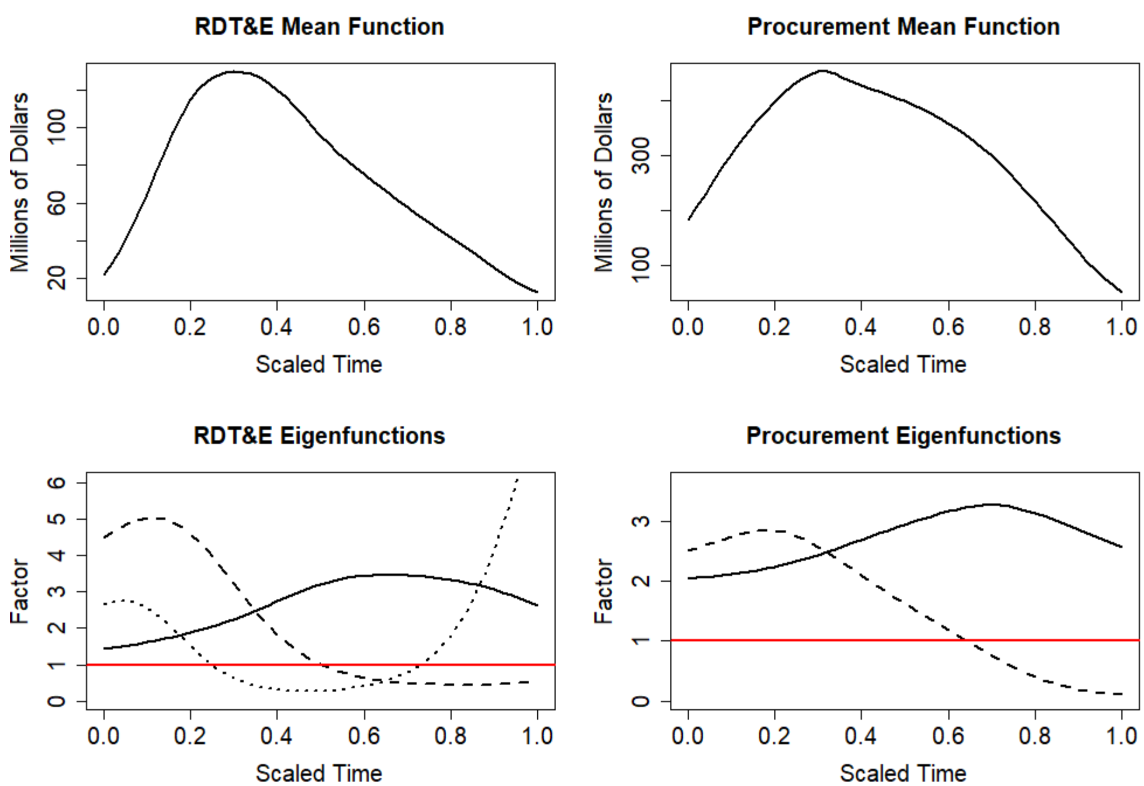

Figure 2. Mean Shape and Eigenfunctions for RDT&E and Procurement

The FPCA process was fit to logged spending profiles; these curves have been transformed back into dollar units. The first principal eigenfunction is represented by the solid line, the second principal eigenfunction is represented by the dashed line, and the third principal eigenfunction is represented by the dotted line. Since eigenfunctions have been exponentiated, these represent multiplicative deviations from the mean. If the eigenfunction is greater than 1, it induces a positive deviation from the mean; if it is less than 1, it is a negative deviation from the mean. The solid red line at 1 represents no deviation from the mean response. The scale of the deviation is determined by the FPCA scores. The FPCA scores are real-valued; thus, if a score is below zero, the eigenfunction flips over the red line (but not symmetrically due to the non-linear transformation).

Identifying Potential Predictor Variables

Given choices for functional forms, the next challenge is to somehow characterize how the distribution of possible actual outcome profiles could be derived for a given initial plan. It seems obvious that different kinds of programs involve different levels of cost risk. There is a substantial literature attempting to identify specific factors that are correlated with program cost and schedule growth. Some factors that have been found by past researchers to be correlated with (unit) cost growth and/or total program cost growth risk include:

- Commodity type (e.g., helicopter, satellite, MAIS, missile, or submarine; Arena et al., 2006; Drezner et al., 1993; Tyson, Harmon, & Utech, 1994)
- Acquiring Service (Army, Navy, Air Force, Joint, Department of Energy) (Drezner & Smith, 1990; Jessup & Williams, 2015; Light et al., 2017; McNicol, 2004)
- New design vs. modification of existing design (Arena et al., 2006; Coonce et al., 2010; Drezner et al., 1993; Jimenez et al., 2016; Marshall & Meckling, 1959)
- New build vs. remanufacture of existing units (Tyson et al., 1989)
- Budget climate at Milestone B (Asher & Maggelet, 1984; McNicol, 2017)
- Number of years of spending prior to Milestone B (Jimenez et al., 2016; Light et al., 2017)
- Schedule optimism (Arena et al., 2006; Asher & Maggelet, 1984; Glennan et al., 1993; Tate, 2016)
- Technology maturity of the program (Adoko, Mazzuchi, & Sarkani, 2015; GAO, 2006)
- Investment size (Bliss, 1991; Creedy, Skitmore, & Wong, 2010)

Because we are not attempting to diagnose causes of cost growth, but are instead only trying to understand and characterize risk (on the assumption that the past is a reasonable guide to the future), we do not distinguish here among risks arising from discretionary choices, environmental factors, or intrinsic program features.

Describing Changes in Cost and Schedule as Changes in Profile Functions

We can model the change in an initial or midlife profile to a final profile by modeling the change in FPCA scores, total spending, and total duration. We saw above that we can model development costs or production costs as being generated from a basis of eigenfunctions. Define ω_{ijk}^{0} to be the FPCA scores of the planned or midlife programs and ω_{ik}^1 to be the FPCA scores of the completed program. Define T_{ij}^0 to be the planned total duration until program completion and T_i^1 to be the total duration when program i is actually completed. Define C_{ij}^0 to be the total cost of an initial or midlife program and C_i^1 to be the total cost of a completed program. Then $\theta_{ij}^0 = (\omega_{ij1}^0, \ldots, \omega_{ijK}^0, T_{ij}^0, C_{ij}^0)$ fully characterizes the cost profile of an initial or midlife program and $\theta_{ij}^1 = (\omega_{i1}^1, \ldots, \omega_{iK}^1, T_i^1, C_i^1)$ fully characterizes a final cost profile. We estimate the conditional (joint) distribution of θ_{ij}^1 given the appropriate program and environmental attributes and the fact that the program's previous estimate was best fit by the FPCA expansion.

There are several possible approaches to this and many choices of how to parametrize the family of curves being fit, but the general method will be the same in all cases. We estimate the distribution of θ_i^1 as a function of parameters θ_{ij}^0 and the historical program characteristics X_{ij}:

$$\theta_i^1 = \left(X_{ij}, \theta_{ij}^0\right)\beta + \eta_{ij},$$

where X_{ij} includes factors such as current estimated cost, Service, budget climate, and so forth, and η_{ij} are independent and identically distributed draws from a multivariate normal distribution centered at the vector 0 with covariance Σ. The vector X_{ij} gives the values of the predictors for historical program i in year j. The vector θ_{ij}^0 contains the elements that describe cost profile for program i in year j. The vector (X_{ij}, θ_{ij}^0) denotes the component-wise concatenation of θ_{ij}^0 onto X_{ij}.

This linear regression model implies a functional fit and distribution over the annual cost profile function $C_{ij}(t)$. Rather than attempting to predict eventual actual cost as a function of initial estimated cost and other predictors, we instead attempt to predict the distribution of the parameters of a function that *generates* eventual cost, given program-specific attributes and the parameters that generate the initial estimate. Note that this is a multiple output regression—we are simultaneously estimating all of the best-fit parameters θ_i^1 and the covariance matrix that describes how those parameters are correlated.

We use a Bayesian estimation framework, starting with a weakly informative prior distribution $F_{prior}\left(\theta_i^1\right)$ and using Markov Chain Monte Carlo (MCMC) estimation to derive a posterior distribution $F_{posterior}\left(\theta_i^1\right)$, including the covariance matrix (Chib & Greenberg, 1995). We do this separately for RDT&E costs and procurement costs, using different families of profile-generating functions, treating their changes in shape and size as independent. Treating development and procurement jointly is a potential area for future research.

Regression Methodology

We compiled original and midlife estimates and actual outcomes for $I = 1{,}278$ historical profiles for RDT&E and $I = 828$ for procurement. For each historical program i at year t, we fit scaled, FPCA scores to the full current profiles (including completed programs):

$$C_{ij}(t) = \mu(t) + \sum_{k=1}^{K} \omega_{ijk}\xi_k(t) + \epsilon_{ij}(t).$$

Let $\theta_{ij} = (\omega_{ij1}, \dots, \omega_{ijK},, C_{ij}, T_{ij}^{p})$ be the parameters of those best-fit curves. Then θ_{ij}^{0} are the best fit parameters to the initial and midlife profiles and θ_{i}^{1} are the best fit parameters to the actual completed profiles. We further decompose C_{ij}^{0} into $(C_{ij}^{0a}, C_{ij}^{0p})$ to be the actual spending that has already occurred (if any) and the planned spending yet to occur. We model the distribution of θ_{i}^{1} as a function of θ_{ij}^{0} and a set of predictor variables X_{ij} simultaneously over all programs, where X_{ij} includes the program-specific and environmental factors previously listed. Parametric linear models are simultaneously fit to obtain a predictive model for the final profile parameters θ_{i}^{1}. The models are as follows:

$$\omega_{\mathrm{i1}}^{1} = (X_{ij}; \theta_{ij}^{0})\beta_{\omega_1} + \eta_{\omega_1},$$

$$\vdots$$

$$\omega_{\mathrm{iK}}^{1} = (X_{ij}; \theta_{ij}^{0})\beta_{\omega_{\mathrm{K}}} + \eta_{\omega_{\mathrm{K}}},$$

$$\sqrt{\mathrm{C}_{\mathrm{i}}^{1}} = \left(X_{ij}; \theta_{ij}^{0}\right)\beta_C + \eta_C\ ,$$

$$\log(\mathrm{T}_{\mathrm{i}}^{1}) = (X_{ij}; \theta_{ij}^{0})\beta_T + \eta_T,$$

where the error terms $(\eta_{\omega_1}, \dots, \eta_{\omega_K}, \eta_C, \eta_T)$ are assumed to be jointly normally distributed. The covariates X_{ij} include information about previously finished programs that had initial planned spending profiles and actual final profiles. Using these historical data, the model is fit to predict final actual profiles using only information available from a program's Milestone B date. The parameters $\beta = (\beta_{\omega_1},, \beta_{\omega_K}, \beta_C, \beta_T)$ are jointly estimated using a Bayesian Seemingly Unrelated Regressions model with prior distributions on the parameters $E\left[\theta_{i}^{1}|X_{ij}\right] \equiv \beta$ and $Var\left[\theta_{i}^{1}|X_{ij}\right] \equiv \Sigma$.

The prior for β has a multivariate normal distribution, calibrated such that prior belief is that there is no change in the profile from the current estimate to final actual profile and no other traits of the initial profile are predictive of the final actual profile. This prior belief is fairly strong in order to induce regularization. This prior choice balances the bias-vs.-variance tradeoff to produce better out-of-sample predictions.

The prior for Σ has an inverse Wishart distribution, chosen such that the equations are uncorrelated and the prior variance is 1.

The joint posterior distribution of β and Σ incorporates the prior beliefs and the historical data to arrive at an updated posterior belief. The Bayesian machinery is especially useful for our purposes because it allows us to obtain random draws from the posterior distribution of β and Σ, which in turn allows us to generate random draws of a final profile distribution $\hat{\theta}_{i}^{1}$ for any program with known initial profile characterized by covariates X_{ij} and θ_{ij}^{0}. This lets us estimate the complete (posterior) distribution of final profiles, rather than just a point estimate and variance measure. We sample from the posterior with an MCMC Gibbs algorithm from Rossi, Allenby, and McCulloch (2005). We draw 400,000 MCMC samples, keep every fourth draw, and discard the first 1,000, leaving us with a Monte Carlo sample of 99,000 draws. Keeping every fourth draw is called "thinning" and reduces the MCMC autocorrelation; discarding the first 1,000 draws is called "burn-in" and ensures we utilize draws after the MCMC algorithm has converged to the posterior distribution.

Regression Data

The data for the regression are the initial estimate and final actual cost profiles for completed historical MDAPs. The earliest program in the data set passed Milestone B in 1982. The data are taken from SARs, together with compiled attributes and environmental factors (as enumerated above) for each program. We apply this method to both development (RDT&E) cost risk and procurement cost risk models, which differ only in which predictor values are used, the number of eigenvectors fitted, and the eigenvector shapes resulting from the estimation.

The following are the specific predictor variables used in this paper:

- ω_{ij1}—the first FPCA score
- ω_{ij2}—the second FPCA score
- ω_{ij3}—the third FPCA score
- $\sqrt{C_{ij}^{a}} = \sqrt{\sum_{t=0}^{j} C_{ij}(t)}$—square root of actual spending
- $\sqrt{C_{ij}^{p}} = \sqrt{\sum_{t=j+1}^{\infty} C_{ij}(t)}$—square root of planned spending
- $\log(T_{ij}^{0})$—natural log of the planned number of future spending years
- The Service overseeing the program (Navy, DoD, Air Force, Army, Department of Energy)
- A commodity type (Air; Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance [C4ISR]; Ground; Ordnance; Sea; Space; other)[1]
- A measure of relative Service budget tightness compared to two years ago
- A measure of relative Service budget tightness over the last 10 years
- A measure of budget optimism—planned spending divided by the mean historical actual spending for this commodity type
- A measure of schedule optimism—planned duration divided by the mean historical actual duration for this commodity type
- Whether the program is based on a modification of a preexisting design (binary)

The measures of relative budget tightness were based on the year the program passed Milestone II/B. The measures of budget and schedule optimism reflect the

[1] More precise commodity categories—e.g., distinguishing helicopters from fixed-wing aircraft—might be useful, given enough data. We found that increasing the sample size in each category led to better results than increasing the precision of the categories.

empirical observation that the average behavior of programs in a given commodity class is a better predictor of cost and schedule than early cost and schedule estimates of individual programs in that class.

Monte Carlo Risk Analysis

General Approach

Suppose that we have budgeted a program at some level, possibly different from its predicted cost profile. Let $B(t)$ be the budgeted funds in year t, and let $C_0(t)$ be the predicted cost that will be incurred in year t. There are many questions we might wish to ask about the program's affordability risk:

- In how many years will the program exceed the planned budget?
- How many total dollars over budget will the program spend?
- What is the probability of exceeding the budget at least once over the FYDP?
- How much contingency funding would be needed to achieve 90% confidence of staying within budget, depending on whether unspent contingency carries over to the next year?

These are all questions of potential interest to both program managers and resource managers. Using the posterior final profile distribution derived from the original profile C_0, we can perform many counterfactual Monte Carlo analyses to answer these kinds of questions. The general pattern for these analyses is as follows:

1. Given the initial development estimate for a program …
2. Define a yearly budget level $B(t)$, and a contingency fund size (if any).
3. Use the regression described above to determine the posterior distribution on the parameters of the best fit to the final actual development profile for the program.
4. Define outcomes or events of interest—e.g., exceeding the budget in some year, or staying within the budget through the entire FYDP, or having planned funds at least as large as spent funds in year 7.
5. For s = 1,…,S (indexing over iterations of the Monte Carlo algorithm):
   a. "Draw" random parameter vector $\theta^{1(s)}$from the posterior distribution.
   b. Compute the corresponding yearly values by evaluating the best fit curve at t = 1, …, $T^{1(s)}$ and computing $\exp(C^{1(s)})\frac{\exp\left(\mu(t)+\sum_{k=1}^{3}\omega_{ik}^{1(s)}\xi(\mathrm{t})\right)}{\sum_{t=1}^{T^{1(s)}}\exp\left(\mu(t)+\sum_{k=1}^{3}\omega_{ik}^{1(s)}\xi(\mathrm{t})\right)}$.
   c. Evaluate and store any events or outcomes of interest.

Note that the value of $T^{1(s)}$ used in step 5b is determined as part of $\theta^{1(s)}$ in step 5a.

After S iterations, calculate the statistics of interest over the stored events or outcomes. For example, count the number of times N that $C^{1(s)}(t) < B(t)$ for t = 1…5, and compute $\frac{N}{S}$. This is the estimated probability of staying within budget for the first five years. The Monte Carlo framework can also allow comparison of different management policies. For example, one could compare the effect of pre-allocating contingency to specific program years, versus maintaining a contingency fund to be spent down over time as needed.

In general, we would do this not only for development profiles, but also for procurement spending. In that case, policy makers might be interested in how much difference it would make to be able to manage both RDT&E and procurement using a single combined budget and/or a single program contingency fund, rather than having to manage separate budgets and contingency amounts due to “color of money” prescriptions.

Figure 3 shows an example of applying this method to a new program, using actual RDT&E cost estimates for a current MDAP. The vertical red line marks the date of the analysis, which is the boundary between past actual funding and projected future funding. The green line shows actual past annual funding; the dashed blue line shows the program estimate of annual future funding. The solid black line is the mean projected funding derived from the FPCA Monte Carlo methodology; the dashed lines mark upper and lower 10% prediction interval bounds[2] year by year as determined using the weighted Monte Carlo.

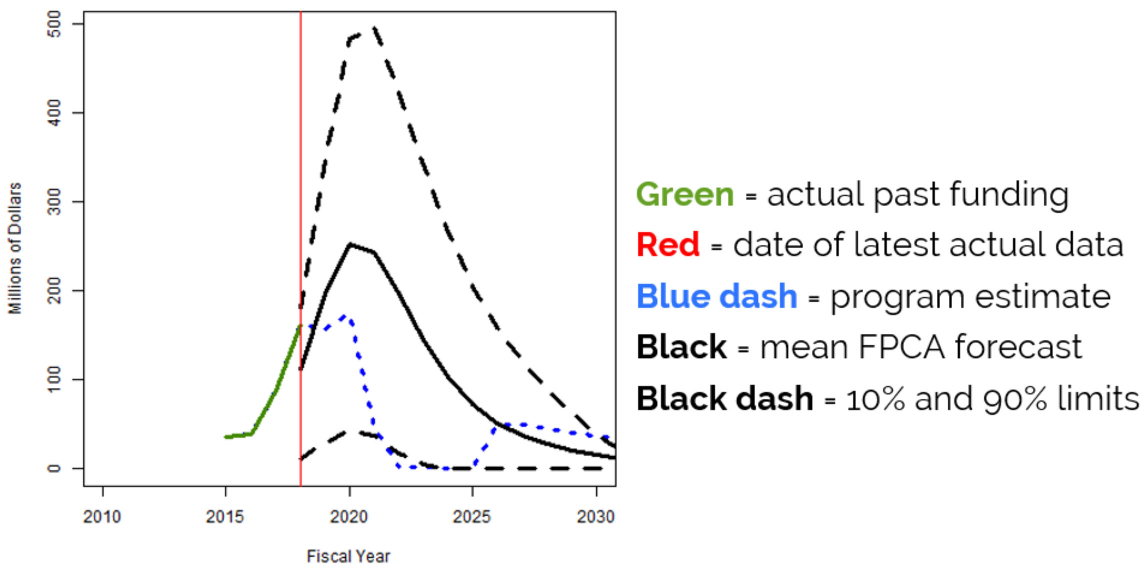

Figure 3. Predicted Future RDT&E Spending for a New Program

Midlife Programs

The method we described in Tate et al. (2018) applies specifically to programs that are just beginning, using an estimate of their future spending profiles. In practice, however, most acquisition programs in any given year are already partially complete, and part of their realized spending profiles is known. We need a method that accounts

[2] That is, the 10% and 90% quantiles of the estimated distribution of possible funding in each year.

for those actual costs to date, and generates profile distributions for the future that are conditioned on that history.

Statistically, it would be difficult to perform conditional FPCA regressions to predict the remaining profiles for RDT&E and procurement, taking the actual costs to date as input factors. Not only would the power of the regressions be greatly reduced (due to paucity of historical programs with a specific cost history), but also the characterization of shapes of actual spending would be at least as complicated as for overall profiles. In the worst case, we would need a separate set of FPCA eigenfunctions for programs with one year of actuals, programs with two years of actuals, and so forth. Fortunately, the Monte Carlo framework for generating posterior empirical distributions provides an alternative that is both computationally efficient and effective.

We implement this method as follows. The original Monte Carlo method as described in Tate et al. (2018) weighs all random draws from the posterior distribution equally, in order to produce year-by-year empirical spending distributions for the program. We modify the method for midlife programs using unequal weighting of these random draws. Instead of weighing all draws equally when estimating the distribution of future profiles, we instead give higher weight to those draws that more closely match the observed history of the program to date. This is comparable to a Nadaraya-Watson estimator (Nadaraya, 1964), and has the effect of conditioning the Monte Carlo–based future estimates on the observed history. The exact weighting scheme can be adjusted to balance between computational efficiency and strict enforcement of the conditioning on the past.

Figure 4 shows an example of this method, applied to a different actual MDAP. Note how the model predicts mean funding levels below the planned level, but with significant uncertainty (including some chance of program cancellation).

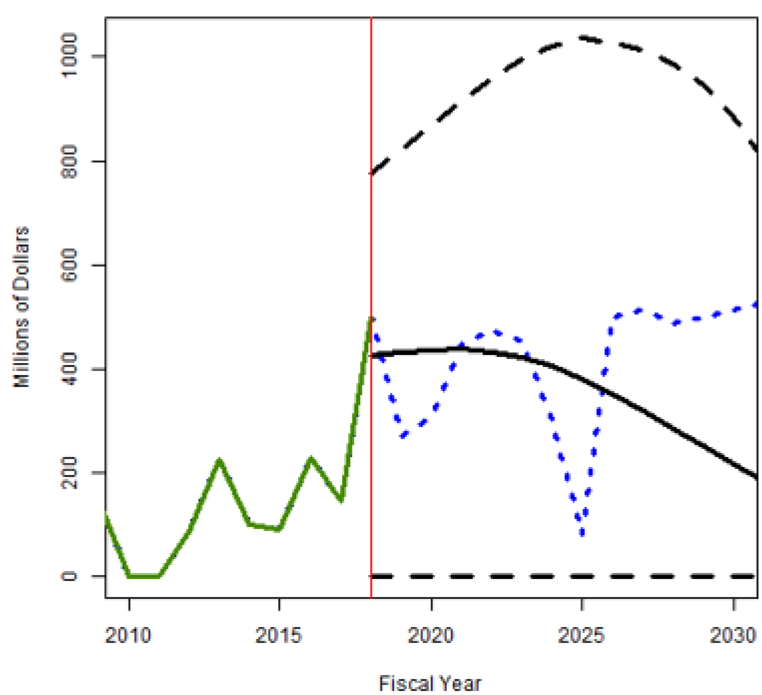

Figure 4. Prediction Intervals for Future Funding of an Ongoing Program

Portfolios: More Than One Program at a Time

We have shown how our model can characterize the affordability risk of a single program's development budget. We noted in Tate et al. (2018) that it would be even more useful to be able to characterize the affordability risk of a group of projects or programs being managed with a common contingency pool. If the conditional outcomes of these programs were approximately independent, this would not be much more complicated than the single-program case. In practice, we know that funding levels among programs within a portfolio are negatively correlated; this is a potential area for future research. For the moment, we treat programs as if they were independent, and incorporate current funding tightness as a predictor of outcomes. This does not affect the mean outcome for each program but does increase the variance.

If we have estimated the $F_{posterior}(\theta^1)$ distributions for each of a set of programs, we can apply the same kind of Monte Carlo analysis to the sum of their annual costs, compared against a collective portfolio budget and contingency fund. This could be done separately for RDT&E and procurement, each with its own budget, or it could be done using a combined investment budget. This would enable true affordability analysis of portfolios as envisioned by the Better Buying Power initiatives,[3] but with considerably more realism than current affordability analyses that are based on point-estimate cost profiles assuming fixed program content and quantities.

One potential use of such a model would be to quantify the benefits of portfolio-level contingency funding versus program-level contingency funding. It is well known in the project management world that allocating reserve funds to specific cost areas before you actually know where the cost growth is going to occur leads to less efficient use of those reserve funds. However, it has historically been difficult to protect funds that are not part of the base budget for some cost element. In the DoD, apart from a highly limited ability to reprogram funds from one program element or line item to another, there is currently no ability to reserve funds for contingency use outside of a specific program's budget. The recent report of the Section 809 Panel specifically recommended expanding the ability of the DoD to reprogram funds across programs and manage contingency at the portfolio level. This research provides some analytical support for those recommendations.

Potential Criticisms of the Method

We noted in Tate et al. (2018) that the utility of these methods assumes, among other things, that historical patterns of cost growth and schedule stretch will persist into the future. This is a conservative assumption, given that observed patterns of cost and schedule growth in major programs have persisted across multiple acquisition systems and regulatory regimes over the past decades. A more nuanced concern is that if resource managers were to actually use these methods to manage portfolios of programs more efficiently, the resulting changes in program outcomes ought to invalidate the models, at least until a new collection of historical outcomes under the new regime could be assembled.

---

[3] Department of Defense, *Better Buying Power*, http://bbp.dau.mil/index.html.

We also noted that these methods offer no insights into *why* costs and schedules deviate from their original estimates (or how this could be "fixed"), and that these methods explicitly model how much funding a program *will receive* in a given year—not how much it needs, or ought to receive, or would receive if there were more money to go around. As such, the model data incorporate the history of negotiations between the Services, the Office of the Secretary of Defense, and the Congress regarding how much to fund programs year by year, and when to cancel them. If there were to be a fundamental change in the dynamic of how those decisions are made, then that, too, might invalidate the link between historical outcomes and future program outcomes, at least until enough new data could be collected.

Finally, we note that the current portfolio modeling approach treats individual program funding levels as independent draws from their respective posterior distributions. This is known to be a weak assumption and is a potential area for future research.

Conclusions

Quantifying Annual Resource Risks for a Program or Portfolio

We have developed a methodology to characterize the year-by-year budget risk associated with a major acquisition program. This methodology can be applied to both development costs and procurement costs, and can be extended to understand the aggregate affordability risk of portfolios of programs. The method allows resource managers to estimate annual budget risk levels, required contingency amounts to achieve a specified probability of staying within a given budget, and a host of other relevant risk metrics for programs. It also allows policy makers to predict the impact on program affordability of proposed changes in how contingency funds are managed.

Research Program Status

The switch from functional regression using parametric curve families to FPCA using nonparametric eigenfunction kernels has significantly improved both the fidelity of the curve fits and the flexibility of the predictive aspects of our approach. We have established that FPCA methods can accurately reproduce historical funding profile shapes for both RDT&E and procurement profiles, and that it is possible to characterize the uncertainty in future spending profiles using the outputs of FPCA and weighted Monte Carlo techniques to sample from the distribution of overall funding profiles while accounting for actual program history to date. This represents a significant improvement in the state of the art; we are not aware of any other technique that has been proposed that can predict time-phased cost and schedule growth distributions for any kind of defense acquisition program, much less a general approach that potentially can be applied to all programs.

Future Research

This technique is currently in the prototype stage and is based on a relatively sparse set of historical program outcome data. There is still much work to be done on establishing the ideal number of eigenfunctions to use in fitting initial and final RDT&E and procurement profiles (respectively), characterizing the distribution of residuals around the best-fit functional curve and utilizing a more flexible mean function in the SUR regression. Additionally, we would like to assess the model's predictive power with an out-of-sample prediction exercise. However, measures for out-of-sample predictive accuracy of functional distributions is an unexplored topic in statistical methodology. There is also a great deal to be learned about how managers could best use the information provided by this method to manage actual programs and portfolios, and what the implications might be for recommending policy changes to acquisition law and regulations. As noted in the Portfolios: More Than One Program at a Time and the Potential Criticisms of the Method sections, the current portfolio modeling approach treats program outcomes as independent. It would be useful to extend this approach to account for correlations among funding levels within a portfolio, or to explicitly model priorities among programs.

References

Adoko, M. T., Mazzuchi, T. A., & Sarkani, S. (2015). Developing a cost overrun predictive model for complex systems development projects. *Project Management Journal*, *46*(6), 111–125. doi:10.1002/pmj21545

Anderson, T. P., & Cherwonik, J. (1997, Summer). Cost estimating risk and cost estimating uncertainty guidelines. *Acquisition Review Quarterly*, 339–348. Retrieved from www.dtic.mil/get-tr-doc/pdf?AD=ADA487532

Arena, M. V., Leonard, R. S., Murray, S. E., & Younossi, O. (2006). *Historical cost growth of completed weapon system programs* (TR-343-AF). Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/content/dam/rand/pubs/technical_reports/2006/RAND_TR343.pdf

Asher, N. J., & Maggelet, T. F. (1984). *On estimating the cost growth of weapon systems* (IDA Paper P-1494). Alexandria, VA: Institute for Defense Analyses.

Bitten, R. E., & Hunt, C. (2017). *Assessing the impact of confidence levels in funding and budgeting NASA science missions*. Paper presented at the ICEAA Professional Development & Training Workshop, Portland, OR.

Bliss, G. (1991, June). *The accuracy of weapon systems cost estimates*. Presented at the 59th Military Operations Research Symposium, U.S. Military Academy, West Point, NY.

Brown, G. E., White, E. D., Ritschel, J. D., & Seibel, M. J. (2015). Time phasing aircraft R&D using the Weibull and Beta distributions. *Journal of Cost Analysis and Parametrics*, *8*(3), 150–164. doi: 10.1080/1941658X.2015.1096219

Chib, S., & Greenberg, E. (1995). Hierarchical analysis of SUR models extensions to correlated serial errors time-varying parameter models. *Journal of Econometrics*, *68*, 339–360. doi:10.1016/0304-4076(94)01653-H

Congressional Budget Office. (2017). *An analysis of the Obama administration's final future defense program* (CBO Report No. 52450). Retrieved from https://www.cbo.gov/system/files/115th-congress-2017-2018/reports/52450-fydp.pdf

Coonce, T., Bitten, R., Hamaker, J., & Hertzfeld, H. (2010). NASA productivity. *Journal of Cost Analysis and Parametrics*, *3*(1), 59–78. doi:10.1080/1941658X.2010.10462228

Creedy, G. D., Skitmore, M., & Wong, J. K. W. (2010). Evaluation of risk factors leading to cost overrun in delivery of highway construction projects. *Journal of Construction Engineering and Management*, *136*(5), 528–537. doi:10.1061/(ASCE)CO.1943-7862.0000160

Drezner, J. A., Jarvaise, J. M., Hess, R. W., Hough, P. G., & Norton, D. (1993). *An analysis of weapon system cost growth*. Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/content/dam/rand/pubs/monograph_reports/2006/MR291.pdf

Drezner, J. A., & Smith, G. K. (1990). *An analysis of weapon system acquisition schedules* (R-3937-ACQ). Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/reports/R3937.html

Glennan, T. K., Bodilly, S. J., Camm, F., Mayer, K. R., & Webb, T. (1993). *Barriers to managing risk in large scale weapons system development programs* (MR-248-AF). Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/monograph_reports/MR248.html

GAO. (2006). *Space acquisitions: DOD needs to take action to address unrealistic initial cost estimates of space systems* (GAO-07-96). Washington, DC: Author. Retrieved from https://www.gao.gov/products/GAO-07-96

Jessup, R., & Williams, J. (2015). *The cost of commonality: Assessing value in joint programs* (MBA professional report). Monterey, CA: Naval Postgraduate School.

Jimenez, C., White, E. D., Brown, G. E., Ritschel, J. D., Lucas, B. M., & Seibel, M. J. (2016). Using pre-Milestone B data to predict schedule duration for defense acquisition programs. *Journal of Cost Analysis and Parametrics*, *9*(2), 112–126. doi:10.1080/1941658X.2016.1201024

Light, T., Leonard, R. S., Pollak, J., Smith, M., & Wallace, A. (2017). *Quantifying cost and schedule uncertainty for major defense acquisition programs (MDAPs)*. Santa Monica, CA: RAND Corporation. doi:10.7249/RR1723

Marshall, A. W., & Meckling, W. H. (1959). *Predictability of the costs, time, and success of development* (P-1821). Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/papers/P1821.html

McNicol, D. L. (2004). *Cost growth in major weapon procurement programs*. Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L. (2017). *Influences on the timing and frequency of cancellations and truncations of major defense acquisition programs* (IDA Paper P-8280). Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L., Gould, B. R., Wu, L., Fuchs, E., & Bronson, P. (2013). *Evidence on the effectiveness of the Defense Acquisition Board process* (IDA Paper P-5011). Alexandria, VA: Institute for Defense Analyses.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, *9*(1), 141–142. doi:10.1137/1109020

Rossi, P. E., Allenby, G. M, & McCulloch, R. E. (2005). *Bayesian statistics and marketing*. Hoboken, NJ: Wiley.

Tate, D. M. (2016). *Acquisition cycle time: Defining the problem* (IDA Document NS D-5762). Alexandria, VA: Institute for Defense Analyses.

Tate, D. M., Coonce, T. J., & Guggisberg, M. R. (2018) Quantifying annual affordability risk of major defense programs (IDA Document NS D-9019). In *Proceedings of* *the 15th Annual Acquisition Research Symposium*. Monterey, CA: Naval Postgraduate School.

Tyson, K. W., Harmon, B. R., & Utech, D. M. (1994). *Understanding cost and schedule growth in acquisition programs* (IDA Paper P-2967). Alexandria, VA: Institute for Defense Analyses.

Tyson, K. W., Nelson, J. R., Om, N. I., & Palmer, P. R. (1989). *Acquiring major systems: Cost and schedule trends and acquisition initiative effectiveness* (IDA Paper P-2201). Alexandria, VA: Institute for Defense Analyses.

Tyson, K. W., Nelson, J. R., & Utech, D. M. (1992). *A perspective on acquisition of NASA space systems* (IDA Document D-1224). Alexandria, VA: Institute for Defense Analyses.

Yao, F., Müller, H., & Wang, J. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, *100*(470), 577–590. doi:10.1198/016214504000001745

Acknowledgments

This work was originally conceived and funded by the congressionally established Section 809 Panel (https://section809panel.org/), and in particular by Commissioners Dr. Allan Burman and Mr. David Ahern. Additional funding and data were provided by the Office of the Secretary of Defense. Dan Cuda and James Bishop of the Institute for Defense Analyses were technical reviewers, and their observations and suggestions improved the paper significantly.

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net