NPS-AM-19-016

ACQUISITION RESEARCH PROGRAM SPONSORED REPORT SERIES

In the Fullness of Time: Towards Realistic Acquisition Schedule Estimates

30 November 2018

Charles Pickar, Senior Lecturer
Dr. Raymond Franck, Researcher
Dr. Gregory Hildebrandt, Researcher
Dr. Bernard Udis, Researcher

Graduate School of Business and Public Policy

Naval Postgraduate School

Approved for public release; distribution is unlimited.

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

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

Abstract

This paper continues a research agenda started in 2016 with an aim of more realistic acquisition program scheduling estimates, especially for the development (SSD) phase. This, our third look at the scheduling problem, starts with a discussion of scheduling data, and how that data could be applied to help the DoD address this challenge. This section includes ideas on how to use acquisition data for the scheduling problem. Next, we present a case study of the result of field interviews with senior DoD leaders. Finally, we present a discussion on using the system performance as a metric.

THIS PAGE LEFT INTENTIONALLY BLANK

About the Authors

Charles Pickar—DBA, is a member of the NPS faculty where he teaches project management, defense acquisition, and systems engineering. Before joining NPS, he led the Applied Systems Engineering Program Area at the Johns Hopkins University Applied Physics Laboratory. He is a retired Army officer with extensive experience in the U.S. defense industry, including director and VP levels at Lockheed Martin, Northrop Grumman, and SAIC. He is the current chair of the Systems Education Technical Committee of the IEEE Systems Council. His research and published work focus on applying systems engineering and system dynamics analytical approaches to defense acquisition problems. [ckpickar@nps.edu]

Raymond Franck—PhD, retired from the faculty of the Graduate School of Business & Public Policy (GSBPP), Naval Postgraduate School (NPS) in 2012. He retired from the Air Force in 2000 in the grade of Brigadier General. His active-duty career included a number of operational tours and staff positions, and head of the Department of Economics and Geography, U.S. Air Force Academy. His published work includes a number of journal articles and research reports in military innovation and defense acquisition management. [cfranck215@aol.com]

Gregory Hildebrandt—PhD, has had an Air Force career including assignments as an acquisition officer, U.S. Air Force Academy faculty member, plus assignments at the Central Intelligence Agency and Office of the Secretary of Defense. Following his Air Force retirement, he has continued service with the RAND Corporation and NPS. His published work includes a number of journal articles in defense economics and RAND reports on acquisition issues. [ggh324@gmail.com]

Bernard Udis—PhD, is a Professor Emeritus of Economics at the University of Colorado at Boulder. He has also served as at the U.S. Air Force Academy and the U.S. Arms Control & Disarmament Agency. His NATO research fellowship examined the costs and benefits of offsets in defense trade. A recognized authority on the economics of defense, his published work includes three books, plus a number of book chapters and journal articles. [bernard.udis@colorado.edu]

NPS-AM-19-016

ACQUISITION RESEARCH PROGRAM
SPONSORED REPORT SERIES

In the Fullness of Time:
Towards Realistic Acquisition Schedule Estimates

30 November 2018

Charles Pickar, Senior Lecturer
Dr. Raymond Franck, Researcher
Dr. Gregory Hildebrandt, Researcher
Dr. Bernard Udis, Researcher

Graduate School of Business and Public Policy

Naval Postgraduate School

THIS PAGE LEFT INTENTIONALLY BLANK

Table of Contents

THIS PAGE LEFT INTENTIONALLY BLANK

Introduction

Weapons system development projects are infamous for exceeding time and cost limitations. Often the reaction to this notoriety is changes at the policy level of acquisition. However, the problem may well lie somewhere else. This paper, like the two preceding papers in this series, suggests we may well be "'*lookin' … in all the wrong places*'" (to paraphrase an old country song)[1] for the causes, because the causes may well lie inside the project and therefore may not be readily addressed by policy changes.

While cost, performance, and schedule are critical variables in any acquisition program, Congress, the media, and policy-makers generally focus on cost, with little attention devoted to the issues of schedule. Moreover, although the DoD has engaged in significant efforts to develop methods for realistic acquisition cost estimates, it has paid considerably less attention to schedules—their estimates and execution. To emphasize the challenge of schedules, Figure 1 provides a macro-level view of the schedule problem. Over the past 20 years, Major Defense Acquisition Programs (MDAP), as reported in Selected Acquisition Reports (SAR), averaged schedule overruns of more than 24 months. Schedule overruns occur for many reasons, and this study examines some of those reasons.

[1] From the words to a song written by Wanda Mallette, Bob Morrison, and Patti Ryan, and recorded by American country music singer Johnny Lee in June 1980.

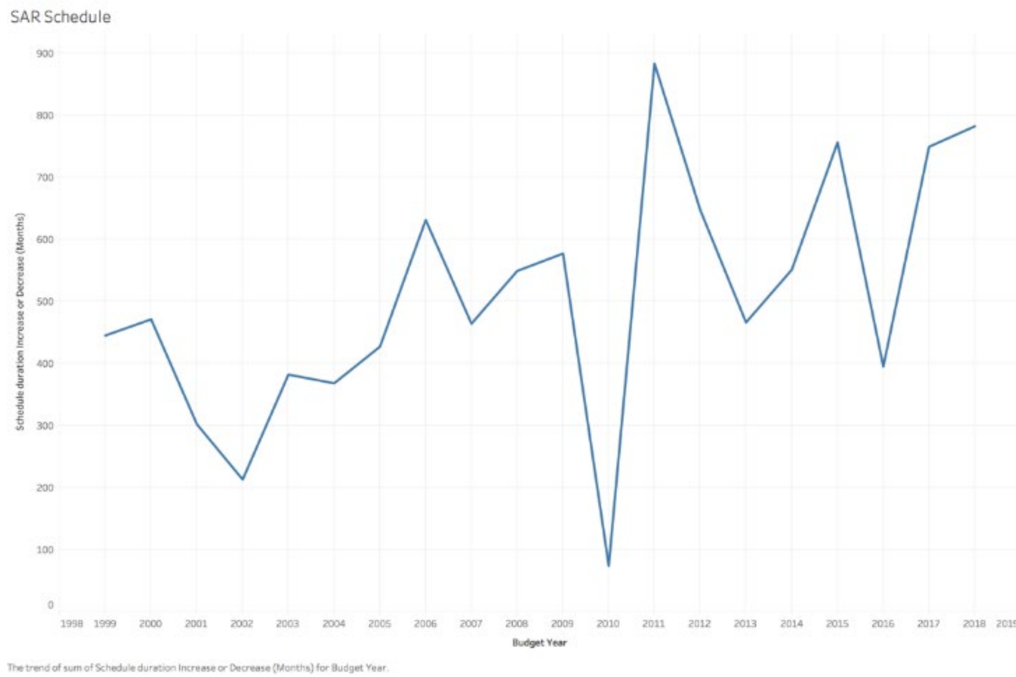

Figure 1. Sum of Schedule Overruns 1998–2017 (Months)

We use a multifaceted approach to examine weapons systems development scheduling to assess the current state and contributing causes of schedule estimating methodologies and suggest different ways to accomplish this difficult process. The overarching research question is as follows:

> What analytical techniques and approaches can be applied to schedule development/analysis to increase the efficiency and effectiveness of schedule estimating and execution?

As long ago as 1988, Morris and Hough were critical of the practice of project management:

> Curiously, despite the enormous attention project management and analysis have received over the years, the track record of projects is fundamentally poor, particularly for the larger and more difficult ones. Overruns are common. Many projects appear as failures, particularly in the public view. Projects are often completed late or over budget, do not perform in the way expected, involve severe strain on participating institutions, or are cancelled prior to their completion after the expenditure of considerable sums of money. p.4

In fact, project management in general, and DoD project management in particular, has been dealing with these problems described by Morris and Hough for decades. We hope to inform these problems because “when problems persist, *practitioners and scholars are getting something wrong* [emphasis added]” (Christensen & Bartman, 2016).

This paper is the third in a series of investigations into alternatives to the way we do schedule estimation today and builds on the research agenda proposed by Franck et al. in 2016, and furthered in Franck et al. in 2017 (Franck, Hildebrandt, & Udis, 2017; Franck, Hildebrandt, Pickar, & Udis, 2017). We start with a discussion of scheduling data, and how that data could be applied to help the DoD address this challenge, and how system dynamics can inform managers of potential schedule problems. Next, we present a case study on the result of field interviews with senior DoD leaders. Finally, we present an exploration of the use of earned value in schedule estimating.

THIS PAGE LEFT INTENTIONALLY BLANK

The Dynamics of Project Management

The concept of time in project management can be divided into two steps, estimating task duration and building the schedule. Both processes require technical expertise and management savvy. First, the technical process of estimating the duration of the project task must be determined. Once duration is established, the management process of project sequencing and scheduling must be defined.

Estimating Activity Duration

Surprisingly, little information is available in the literature on the "how" to estimate the elements of a schedule—the task duration. While the major defense contractors have formal in-company processes, little formal literature is available on the specifics of task estimation. Further, most available information on estimating task duration is found in project management textbooks, but even then, the specifics are scarce.

The PMBOK (Project Management Body of Knowledge) lists five methods for estimating project activity duration. These methods include (Project Management Institute [PMI], 2017):

- Expert Judgment
- Alternatives Analysis
- Published Estimating Data
- Project Management Software
- Bottoms-Up Estimating

Expert judgment acknowledges that technical and engineering experts should be able to estimate the effort necessary to accomplish tasks and translate those estimates to duration. This assumes the chosen experts have significant experience in the execution of those tasks and are therefore competent to judge time required (Hughes, 1996).

Alternatives analysis recognizes that activities or tasks can be accomplished in different ways—alternatives. These different ways include defining different techniques, differing levels of resources, and using different machines.

Published estimates are databanks that gather resources measures. These measures include hourly rates by skill level, acknowledged production rates for various development and manufacturing activities. In most cases, this data is available internal to the organization. However, there are data companies that track and report this data. An example is the IEEE-USA Salary & Benefits Survey. This data is often available for different locations in the United States, as well as worldwide.

Project management software is not really an estimation method. Instead, it provides a means to identify and organize information necessary for resource estimates.

Finally, an engineering or bottoms-up estimate is a comprehensive schedule (and cost) process that starts at the work package level and aggregates costs to build a complete estimate. Bottoms-up estimates are necessary when schedule activities cannot be accurately estimated using another technique. As the name implies, bottoms-up estimates start at a level of activity or task that can be confidently estimated. The activities are then rolled-up to the required level. These estimates are extremely work-intensive but are also the most accurate.

Other recognized methods include parametric techniques. A parametric or top-down estimate builds an activity estimate for the development project from historical data comparing variables through a statistical relationship. All the methods listed are used to estimate the length of time for each of the activities or Work Breakdown Structure tasks lists. "Simply stated, the duration of an activity is the scope of the work (quantity) divided by a measure of productivity" (Hendrickson, Martinelli, & Rehak, 1987, p. 278).

Thus, activity duration estimation establishes the actual time required to complete discrete tasks in an overall project, while project scheduling fixes the start and end dates, as well as execution approaches of the project. Once the overall schedule is established, management activities driven by either time and/or resource constraints will determine the actual execution of the project (Schwindt & Zimmerman, 2015). The analogy that comes to mind is that of an orchestra. The individual instruments (and of course the musicians) are the discrete tasks of the project. The orchestra leader is the project manager. And the music score is “plan” the orchestra leader uses to execute the “project.” Building on this information, the next step in this effort is to identify schedule data that can be used to augment these estimating activities.

Schedule Data

While there is significant information available on DoD procurements, most of that information is on cost. In order to effectively examine project schedules, we must be able to better understand those schedules. It is common knowledge that weapons system development project overrun their schedules. However, we need to be able to determine what causes schedule overruns, as well as an actual measure of the development time.

Data for this research was obtained from the Defense Acquisition Management Information Retrieval (DAMIR) database, a repository for, *inter alia,* the DoD Selected Acquisition Reports (SAR). The SAR is a summary of the acquisition data of selected Major Defense Acquisition Programs (MDAP). Table 1 provides a list of delay factors, as well as maximum and minimum delays, as reported in the SAR during the period 1997–2017.[2]

Table 1 Schedule Delay Factors

Delay Factor	# instances	Maximum Delay (months)	Minimum Delay (months)
Administrative changes to schedule including updates to APB, ADM changes, as well as changes resulting from Nunn-McCurdy processes and program restructuring	460	168	5
Technical	291	60	4
Testing delays	283	66	1
Delay in availability of key capabilities/facilities (launch vehicle/testing facilities/IOT&E units)	3	13	6
Budget/Funding Delays	52	43	1
Delays attributed to the Contractor	50		
Delays because of Rework	16	4	1
External events such as inflation, earthquakes, labor strikes, etc. *(Force Majeure)*	4	4	1
Delays due to Contracting/Contract Negotiation/Award delays	29	27	1
Actuals (updating previously reported dates to actual occurrence)	172	13	-39

[2] The data described are from an unpublished study by the author of the delay factors for DoD program 1997–2017. The study is an initial attempt at quantifying schedule delays in program execution with the intent of using those delays to better inform project planning.

These delay factors suggest that PMs should plan for the time necessary to deal with oversight, information reporting, and both the time takes, as well as the impacts of decisions—internal and external to the program. As the Government Accountability Office (GAO) pointed out in a 2015 study, the program office overheads associated with administrative activities added on an average of two years to complete:

> Programs we surveyed spent on average over 2 years completing the steps necessary to document up to 49 information requirements for their most recent acquisition milestone. This includes the time for the program office to develop the documentation and for various stakeholders to review and approve the documentation.

Figure 2 provides a trend line and forecast of the delays identified. Using this data, the forecast total delay months across all programs in 2019 would be 712 months, and in 2020 that forecast would increase to 729 months.

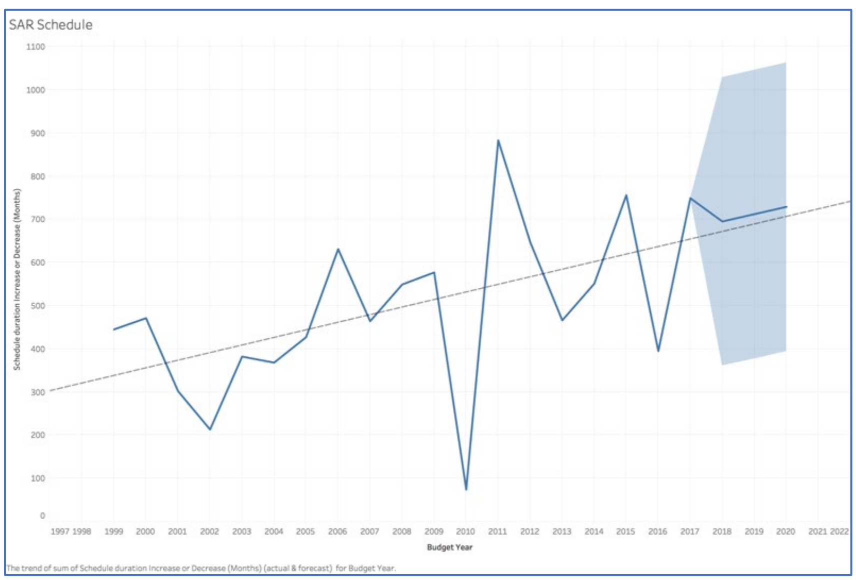

Figure 2. Trend Line Showing Forecasted Schedule Increases

Applying the Data

Our previous paper introduced the rework concept, shown in Figure 3. As noted, the CPM/PERT approach to scheduling precludes the use of data at the program schedule level. And while some companies track task estimation data, that data is often proprietary and more focused on technical process estimation (Godlewski, Lee, & Cooper, 2012).

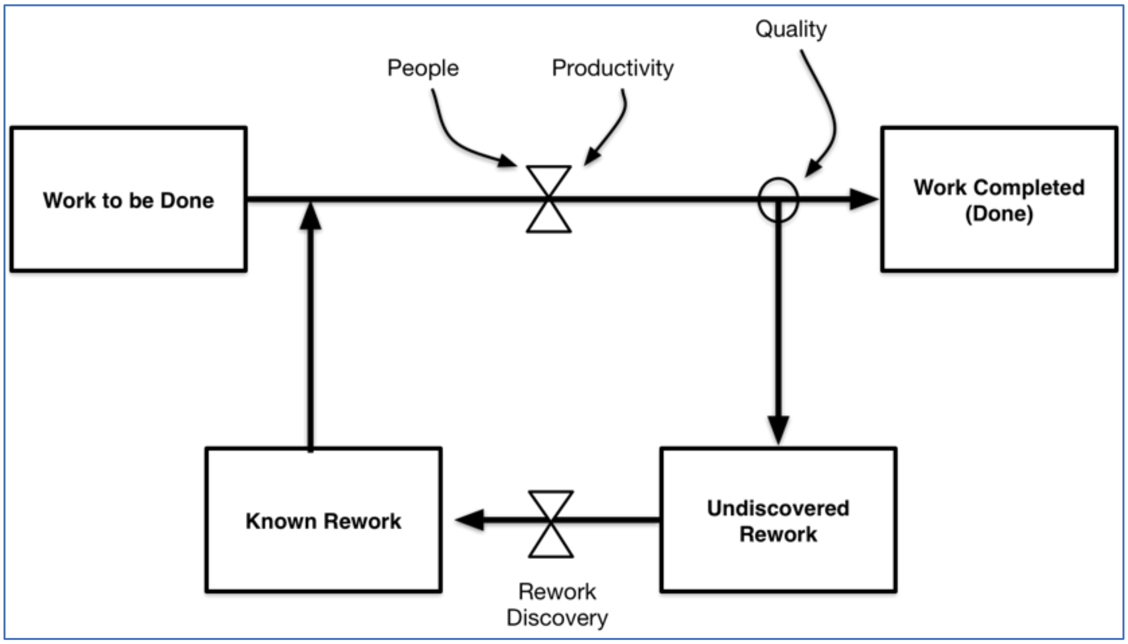

Figure 3. The Rework Cycle

The basic assumption that work proceeds as planned in the network from start to finish is naïve at best (Franck et al., 2017). System dynamics can account for the feedback that results from decisions made in the execution of a project. A project network using CPM/PERT techniques depends on each task being completed in the defined order established. While most PMs attempt to maintain that order, the reality of dynamics intervenes. That reality means that network analysis cannot capture the progress of a project (Williams, Eden, Ackermann, & Tait, 1994).

A tool used in system dynamics to capture cause and effect is a causal map. The causal map becomes a tool used for the development of a model of the delay factors identified. Figure 4 is an initial causal map capturing some of the identified factors in weapons system program schedule delays. The factors shown are a subset of those identified for brevity in this paper.

Delay factors plus the effects of rework, decision wait time, tasks start delay, and other disruptions result in the PM (or PMO) recognizing a schedule problem (delay in the critical path). Invariably, the PM must take action to attempt to return the project to the equilibrium expressed as being on schedule. Thus, the PM could approve overtime, reschedule, or take some other mitigation. The pressure to get back on schedule is driven by many factors including cost considerations, pressure from the oversight organizations, and in weapons systems development, the necessity of delivering capability to the warfighter in the most efficient time. Regardless the reason, the PM "does something." The plus and minus signs indicate the effect of the actions taken.

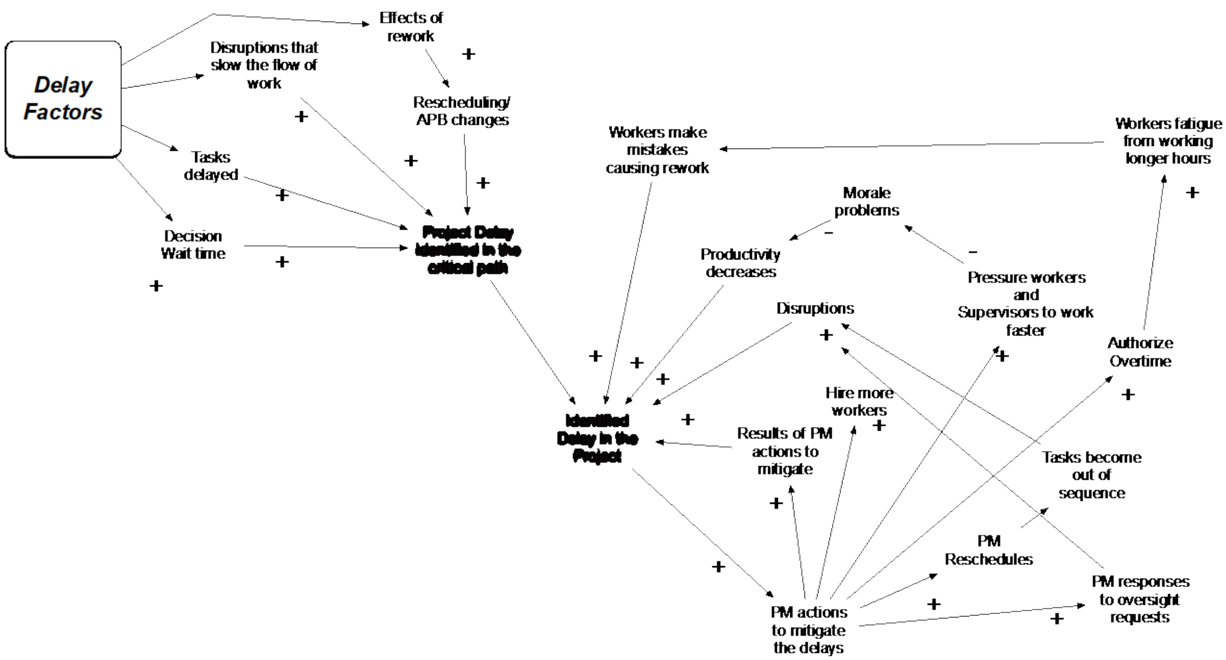

Figure 4. Delay Factors Triggers for Project Delays

(Adapted from Howick, 2003)

A project is a dynamic system with feedback loops and invariably decisions taken to address one problem have impact on or create new problems. For example, approving overtime does initially address schedule issues as more work is being done in shorter periods. However, a recognized problem of overtime is fatigue. Fatigue causes workers to make mistakes, and those mistakes result in having to redo the work, thus perpetuating problems that were thought solved.

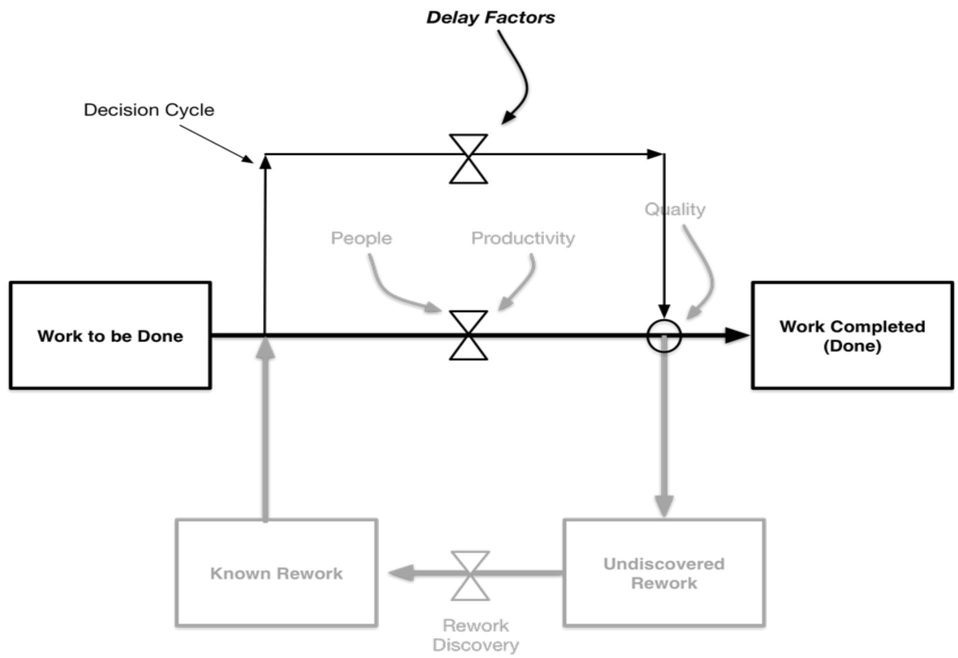

Figure 5. Notional Decision Cycle Added to Rework Cycle Diagram

Similarly, hiring more workers causes more problems. Assuming the new workers have the requisite skills, they need to be trained/acclimated to the actual project situation. In the *Mythical Man Month*, Brooks (1995) explained how this concept works in software development. In reality, it is universal.

Finally, while many of the delay factors identified from the SAR analysis can be explained in Figure 4, others require further examination. One of the biggest challenges is the area of decisions, both internal and external. The internal decisions drive many of the actors discussed above. However, the PM must also deal with external decisions that can eventually impact the development.

Figure 5 is a notional graphic that represents a generic decision cycle in the context of the rework cycle. While the results of this data analysis included rework, the majority of the identified delay factors were decision-focused. Those decision-centric factors included represent this decision cycle. The notation is shown between the work to be done and work completed boxes because many of the decisions identified occur outside the project manager’s purview. The exogenous factors identified cause either reactions to those factors or force other internal decisions. While not normally a part of the rework cycle, we suggest that a formal appreciation of a decision cycle, and the time it takes for decisions to be made both internal and external to the program management cycle must be considered.

Conclusion

This section of the paper presented schedule information gleaned from Selected Acquisition Reports and suggested a model to show how that information can be best understood in the context of the decisions necessary to model weapons system acquisition programs. To be clear, we are not advocating to replace the CPM/PERT methods used today. At best, system dynamics is an adjunct to those methods in use. Instead, we suggest that we should recognize the dynamics at play in any weapons system development, and once recognized use the appropriate tools to better our execution.

No program manager sets out to overrun a schedule. “However, [c]lients increasingly value not only cost and schedule control but cost and schedule certainty” (Godlewski et al., 2012, p. 18). Those clients for defense acquisition products seek certainty as well, both in cost and schedule. It is no secret that current methods for estimating and executing schedule are insufficient. In fact, certainty is one of the potential benefits of this examination of schedule factors. Project certainty starts in effective schedule planning by using the right tools.

F-35 Schedule: A Case Study in Progress[3]

Much has been written about the F-35 program in many venues. These include excellent analyses of cost growth from IDA, RAND, and others (Arnold et al., 2010; Blickstein et al., 2011). These studies focused on costs (as they were charged to do), with accordingly less attention to schedule.

Yet schedule is a significant part of the F-35 story—arguably just as important. Schedule slippages were a significant problem for the F-35 program itself, but also for combat aircraft forces in the United States and elsewhere. As one respected observer put it, “The failure of the so-called fifth-generation fighters ... to arrive on time and on cost is having cascading effects throughout U.S. and allied fighter forces” (Sweetman, 2012). These impacts included capability gaps (e.g., a severe readiness situation with Marine Corps F-18s);[4] more time for rivals to develop counters (and therefore lower F-35 effectiveness after IOC); and costs associated with extending service lives of “legacy aircraft” (Tirpak, 2011).

Defense acquisition professionals know a lot about “what” has happened. “How” and “why” it has happened is less clear. Our last essay (Franck, Hildebrandt, Pickar & Udis, 2017) undertook an inquiry as to the “hows” and “whys” of this case. We asked how a program whose lineage included the Common Affordable Lightweight Fighter (Global Security, 2018) became the F-35, which is not very common (Bogdan, 2012), definitely not lightweight, of still debatable affordability (e.g., GAO, 2017[5]; Capaccio, 2018), and arguably not a fighter (Airpower Australia, 2017).

The publicly-available literature is not terribly enlightening, although a few interesting leads are discernible. We closed with an intent “to learn more in future

---

[3] We are greatly indebted to a highly-placed, well-informed DoD official for many of the insights that underpin this section of our paper. Chatham House Rule applies.

[4] Described recently as a “death spiral” (Friedberg, 2018).

[5] However, the terms of the discussion appear to have shifted from acquisition to sustainment costs.

inquiries" (p. 420). Since then, the field interview method has brought new insights to many aspects of the F-35 program.

Continuing this work in progress, we concentrate on some useful hypotheses we've gleaned—with assessments of them being a matter for further inquiry. These hypotheses[6] concern program management, technology and engineering, and the lure of new technologies. Careful readers will note they are not mutually exclusive and are interrelated in a number of ways.

Program Management

Program management can be characterized as poorly structured from the start with an underequipped and over-burdened program office, which enabled bad decisions.

Program Structure: The program turned out to be well designed to fail. Basically, Lockheed-Martin (LM, the prime contractor) had considerable discretion and control over a highly complex program with a vague set of requirements. Moreover, the incentive structure was not well designed ("poor" according to at least one authority). This was a principal–agent situation (Kreps, 1990, Chap. 16) with the principal (DoD) unable to fully monitor the agent's (LM's) behavior, or to incentivize good results. Also, the IDA Root Cause Analysis noted a likely "lack of clear incentives" in the program (Arnold et al., 2010, esp. p. S-2). One result was a strained relationship between LM and the DoD ("worst I've ever seen"; Bogdan, 2012).

The program strategy reflected a number of optimistic framing assumptions. These included the assumption that joint programs save money, as well as new and promising, but untried, methods expected to significantly reduce risk and time. Perhaps the most optimistic of these basic (framing) assumptions was that joint development of somewhat disparate weapon systems would save money (Lorell, 2013).

---

[6] Although readers will likely not agree with all the details, few, if any, will be surprised.

This latter set included the assumed benefits of recent acquisition reforms and better simulation methods expected to reduce flight testing. All this led to an aggressive schedule involving tight timelines with a high degree of concurrency accepted *a priori* (Blickstein et al., 2011, p. 37).

When these assumptions were not borne out, schedules stretched out, and costs grew. The RAND and IDA Root Cause Analyses, for example, concluded "optimistic cost and schedule estimates" constituted a major cause of program difficulties (Arnold et al., 2010; Blickstein et al., 2011, p. 37). And, there is now an excellent case for common development of combat systems with heterogeneous requirements not saving money (Lorell, 2013, esp. pp. xvii, 31, 32).

Program Office: The F-35's DoD management team was assigned a task that included serious complexities in both technical and management dimensions. Management difficulties included coordination of 11 stakeholders (three U.S. and eight international) with varied operational needs while complying with the U.S. International Trade in Arms Regulations (ITAR) regime.

Cascading effects of program difficulties made the work even more complex. One example was a weight growth problem early in the program (precipitated in part by entering development with a slender weight growth margin). This necessitated a larger engine, which in turn necessitated a major redesign of the fuselage to accommodate that engine, resulting in significant cost growth and schedule delay (Blickstein et al., p. 53).

Basically, the theory of F-35 program success centered on quick development followed by quick transition to high-rate production (Blickstein et al., 2011, p. 43). The path to F-35 success was paved with a series of framing assumptions. Each of them, taken separately, was at least somewhat optimistic. And successfully getting down that path entailed all of them being true. When unexpected difficulties (or problems that were assumed away) emerged in the development process, there were cost and schedule difficulties directly related to that problem. There were also "spillover" problems because of effects on other parts of the design (Blickstein et al., 2011, p. 55). Basically, the acquisition strategy turned out to be something of a house of cards.

Given its highly complex and demanding mission, the F-35 program office was woefully underequipped at crucial junctures. Requirements discipline in the formative period has been characterized as “weak” and unable to deal effectively with several changes internal to the program (e.g., tech insertions, revised development plans) and external (e.g., threat evolution). In addition, there were, at times, significant mismatches between program office needs and personnel skills aboard.

Some tools of program management were inadequate, particularly for schedules. From a program perspective, schedule management tools proved hard to use; not well tied to resource use; insufficiently flexible to account for risk and program perturbations; and not well supported with data from historical experience. As program difficulties arose, there was no credible means available to estimate schedule implications.

These are, of course, difficulties that afflict any defense acquisition program. However, new, complex, difficult, advanced systems like the F-35 suffer more. Another difficulty was rotating new program executive officers (PEOs) every few years. Accordingly, both opportunity and incentive to reorient the program were in very short supply. This particular pattern was broken in 2012 with an indefinite-term PEO.

In addition, as problems continued, the program office was subject to a rather onerous oversight regime, with attendant political pressures and constraints. The one-year F-35B probation period is one example (Franck et al., 2012, esp. pp. 57–59).

Program Execution: Bad Decisions. The factors cited above facilitated bad decisions. The flawed assumptions that underpinned the acquisition strategy did not receive sufficient scrutiny (perhaps related to leadership tenure). In an atmosphere of pervasive optimism, relatively pessimistic assessments (such as the CAIG report in 2001) had little apparent effect on program management (Blickstein et al., 2011, p. 37). Requirements remained in some degree of flux well into the program life, with corresponding effects on program stability.

Heavy reliance on test data from simulations and non-scale airframes led to problems that greatly delayed the test program when those new data sources proved less useful than expected.

The F-35 Helmet Mounted Display (HMD) was a major technical advance with great promise but high risk and no guarantee of success. And a natural programmatic hedge, head-up display (HUD), was cancelled early in the program. This meant that lags in HMD development became a major threat to program success (Bogdan, 2012).

Program office personnel clung closely to a commonality standard among the three models, with cost growth and delays associated with fixing one model's problems among all three models. (This seems to make sense if the F-35 is one unified program. Less so, if there are three programs with commonalities.[7])

Technology and Engineering

The optimism that set the theme for the management strategy also pervaded the technology assumptions. There was a strong proclivity to underestimate the difficulties and risks. While, for example, there was a fair amount of experience with stealthy aircraft designs within the U.S. defense industrial base, the F-35 was nonetheless a major leap forward. As RAND's Root Cause Analysis noted, the basic technical requirements were very demanding. This is illustrated in Table 2. Given the high degree of commonality specified for the F-35, if one model needed to meet certain design objectives, all models needed to do the same. It took considerable ingenuity to design an airplane whose morphology accommodated all these requirements, and the solution that emerged was not robust (Blickstein et al., 2011, esp. p. 37).

[7] Gen. Bogdan (2012) eloquently stated the separate-programs perspective.

Table 2. Required Features for F-35 Design

(Adapted from Blickstein et al., 2011, Table 4.6, p. 49)

	STEALTH	STOVL	SUPERSONIC
Engine Inlets	Small	Large	Specific Shapes
Fuel Capacity	Internal Only	Small	Large
Airframe Shape	Specific (radar signature)	Specific (Weight Distribution)	Specific (speed regime transitions)
Materials	Stealthy airframe skin	Light Skin for vertical landing	Strong Skin (speed regime transitions)

Accordingly, there was little margin for error or unexpected difficulties; one example was 6% allowance for increased weight. That reserve was exceeded early in the program, which necessitated a major redesign exercise (Blickstein et al., 2011, p. 47, 53, as noted above).

Given the demanding nature of the original design and slender margins for error, there was nonetheless a definite willingness to push the technical envelope and take significant risks. Thus, for example, the Helmet Mounted Display (discussed above) was a major technical advance, with a natural hedge (HUD) discarded early.[8]

There was likewise a propensity to trust new and promising, but not fully validated, engineering methods. These included computer simulations substituting much of the testing normally accomplished in the air. The result was a test program generally behind and in a catch-up mode (e.g., DOTE, 2016, esp. p. 31).

The Attraction of New Technologies

Technology insertions occurred with some frequency throughout the development program during both the JAST (Joint Advanced Strike Technologies) and JSF (Joint Strike Fighter) periods. These included the Autonomic Logistics Information System (ALIS) and the Helmet-Mounted Display. ALIS seems to have been regarded as

[8] Reasonable people can disagree as to whether this is a management issue, technical issue, or both.

a logical extension of onboard aircraft diagnostics (Steidle, 1997, p. 9). However, more than a decade later, problems with ALIS were an existential threat to the entire program (Bogdan, 2012).

Likewise, the evolution of the F-35 from an affordable, limited-capability companion for the F-22 (*inter alia*) to a "situational awareness machine" seems to be related to some major advances in sensor capabilities that the F-35 program adopted. International stakeholder interests were also involved: "JAST … was … designed to have the smallest possible sensor suite and be dependent on external information sources. … [But] most of the export countries did not have (those sources) in their inventory," and the F-35 became a battlefield information producer (Keijsper, 2007, p.135).

Such initiatives, taken in isolation, were undoubtedly viewed as sensible at the time. However, the cumulative effect of a series of sensible decisions can be a horrible end result.

The last word on the new technologies and platform performance issues might well come from General Deptula (2016):

> Current systems are largely expected to operate in a semi-autonomous fashion, with a basic level of collaborative engagement with other platforms. *These shortcomings place pressure on individual assets to possess numerous internal capabilities. The complexity inherent to this approach drives lengthy development cycles, which in turn leads to requirement creep, time and cost overruns, and delays in capability* [emphasis added]. (pp. 6–7)

This seems an indirect reference to the F-35 we have developed and produced.

Some Economist's Observations

First is the importance of incentives. As the IDA RCA (Arnold et al., 2010) and informed acquisition professionals have noted, there is reason to question the usefulness of the F-35 program's incentive structure.

Second, and related, is that the F-35 program strategy (especially in earlier stages) looks very much like a bad solution to a principal–agent problem. The agent

(LM) was not well incentivized and not subject to sufficient oversight from the principal (DoD).

Third is a theme from transaction cost economics. After source selection and after the contract signing, the government's power over LM declined significantly. One rule of thumb is that government leverage on industry is at its maximum just before the contract is signed, and before the status changes to bilateral monopoly.[9] This seems to have been, at least in retrospect, insufficiently appreciated when the source-selection contract with LM was signed.

Some Questions for Further Investigation

1. Can an acquisition program schedule become self-stretching? A simplified version of this hypothesis goes something like this. System complexity entails a lengthy development program. Over time, various technical improvements present themselves, some of which are adopted. These technical insertions (even if done well) nonetheless add to system complexity or estimated program schedule (or both). This cycle is summarized in Figure 6.

While this influence diagram seems plausible, the strength of these connections and their total effects on program schedules are subjects for further inquiries.

[9] This is sometimes called the "fundamental transformation" in transaction cost economics literature (e.g., Williamson, 1996).

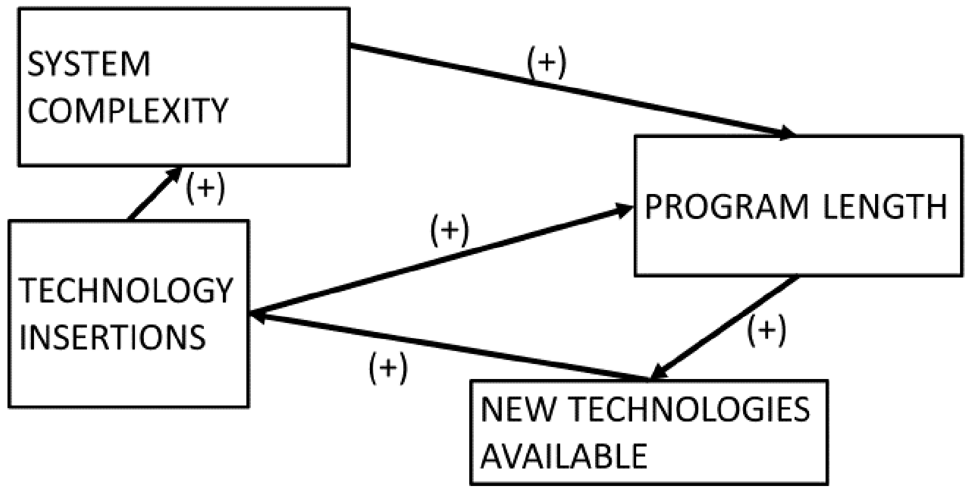

Figure 6. Self-Stretching Acquisition Program Schedules?

2. To what extent do weak schedule estimation and management tools affect program performance? There are excellent reasons to believe that scheduling estimates are sometimes not realistic. What schedule management tools do PMs and POs lack? How can those gaps be addressed?

In some macro sense, program managers make choices among cost, performance, and schedule. And the JCIDS Manual (CJCS, 2015, A-9) sensibly recommends those tradeoffs. PMs have reasonably good cost estimation tools, reasonably good indicators of system performance,[10] but not good ways to estimate schedules, especially if the original program experiences requirements growth.

The acquisition community has professional cost analysts and estimators, but none who deal with schedules as a body of professional practice. Thus, inserting new technologies, which undoubtedly can improve performance, cannot be reliably analyzed with respect to schedule implications. Current scheduling tools are generally recognized

[10] If the performance needs are well specified and accurately embodied in engineering specifications, this statement is reasonably accurate.

as inflexible; hard to use; not well related to resources; and not dealing with program risk.

This question offers perhaps some scope for gap analysis—to be investigated through case studies and interviews with subject matter experts.

3. We’ve received excellent insights into the government perspective of the F-35 program. Can we learn more from our industrial partners?

Analysis of Intra-Program Schedule Growth During Weapon System Acquisition: The Case of the F-35

Introduction

This analysis focuses on intra-program schedule growth using F-35 annual SAR data from December 31, 2001, through December 31, 2017. This analysis can be contrasted with other analyses that have focused on inter-program schedule growth. In Figure 7, we show an example of the trends that have occurred for fighter aircraft from Source Selection to Initial Operational Capability from WWII to the present.

When analysis turns to intra-program growth, initially it includes a discussion of key program dares for the F-35s. Included are Milestone B, Milestone C, and service-specific IOCs. We contrast the projected milestone dates contained in the Dec-01 SAR with the milestone dates reported in the Dec-17 SAR, and show substantial intra-program growth of these key indicators

In addition, we show the important difference between the Acquisition Program Baseline (APB) and the SAR Baseline. These baselines can impact the level of the level of Baseline RDT&E and Procurement and the cost-planning estimates for the two appropriation categories. The APB is used to determine if there has been a Nunn-McCurdy beach, while at the start the SAR Baseline RD&E and Procurement Baselines are estimated and remain in effect throughout a particular Baseline's period.

We also discuss the major program cost variance categories: Quantity, Schedule, Engineering, Estimating, Support and Economic, and explain how both the prior value and current value of these variances combined with the SAR Baseline costs to estimate current program RDR&E and Procurement costs.

Drawing on the discussion of milestones, baselines, and program variances, we next present an empirical forecasting model that explains the tine difference between Milestone C and Milestone B. The key explanatory variables are Prior Total RDT&E Variance, and Prior Total Procurement Variance. As both variables can be computed from the prior SAR report, they should be helpful for forecasting the dependent variable. Also included are a categorical variable that distinguish between RDT&E and Procurement costs, a categorical variable that identifies the two SAR Baseline periods, and the F-35 procurement production quantities.

The SAR reports also include extensive contract data, and we turn next to an explanation of contract Schedule and Cost Variance. An empirical model is presented that links Prior Program Variance to Prior Contract Schedule Variance. Also included in the model are number of days between the Contract Award Date and the SAR date as an additional explanatory variable.

Finally, observations and recommendations are provided.

Inter-Program Analysis

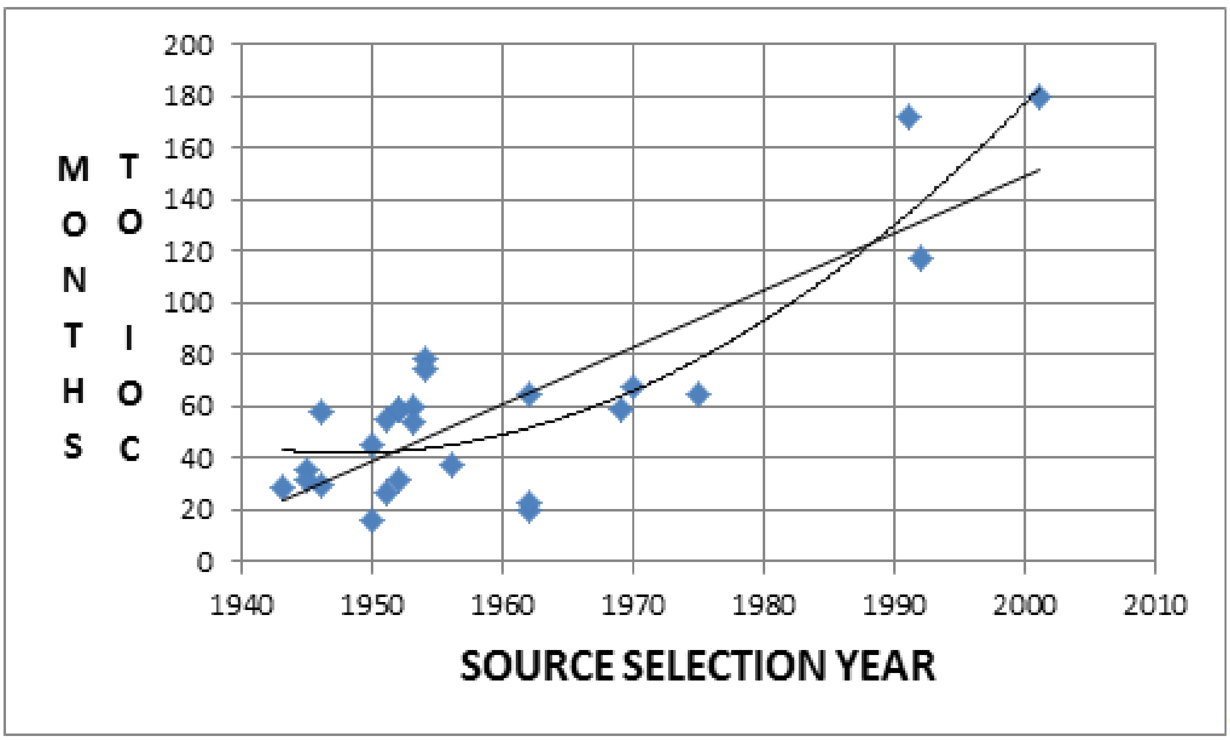

Figure 7. Time Curve for U.S. Fighter Aircraft

(Source: Raymond E. Franck, Jr., Brig Gen, USAF Retired, extended previous inter-temporal schedule duration plots by including WWII period.[i])

Figure 7 provides a scatter plot and two reference lines depicting the Time in Months from Source Selection year for Engineering and Manufacturing Development (EMD) to Initial Operational Capability (IOC) for a wide range of U.S. fighter AC beginning in WWI II. A clear upward trend is shown.

Recently, the DoD has become increasingly concerned with schedule growth. A recent initiative that has been promoted this new emphasis is “Should Schedule.”[ii] A great deal of the focus has been on the Engineering and Manufacturing Development (EMD) period, which has been employed in different analyses as the number of months from Milestone B to IOC, the time period from MS B Contract Award to Milestone C, and similar measures.

The focus of much of this analysis has been on inter-program schedule growth, and the starting point of such analysis can be represented by Figure 8, which displays an increase in the time period during which engineering development occurs.

In contrast, this analysis addresses intra-program schedule growth using the F-35 aircraft. The analysis does not provide a complete picture of the sources of the schedule growth experienced but identifies a number of factors associated with program schedule growth.

F-35 Milestones, Procurement Quantities, and Baselines

Table 3. Key F-35 Milestones and Procurement Quantities
(Source: SARs—December 31, 2001, and December 31, 2017)

Beginning with the Dec-01 SAR, there have been four different Acquisition Program Baselines (APBs). During the same period, there have been two SAR Baselines. Throughout a baseline period, “Objective” RDT&E and Procurement costs remain constant, provided Base Year Dollars do not change. However, beginning with the Dec-11 SAR, Base Year Dollars changed from 2002 to 2012.

RDT&E and Procurement Baseline Costs and Current (then Year) Baseline Cost Estimates

Table 4. Two Baselines Used in System Acquisition

SAR	APB	SAR Baseline	Base Year $
Dec-01	26-Oct-01	26-Oct-01	2002
Dec-03	17-Mar-04	26-Oct-01	2002
Dec-11	26-Mar-12	26-Oct-01	2012
Dec-12	26-Mar-12	26-Mar-12	2012
Dec-14	18-Jun_14	26-Mar-12	2012
Dec-17	18-Jun_14	26-Mar-12	2012

Beginning with the Dec-01 SAR, there have been four different Acquisition Program Baselines (APBs), where the dates indicate when a new baseline is initiated. During the same period, there have been two SAR Baselines. Throughout a baseline period, “Objective” RDT&E and Procurement costs remain constant, provided Base Year Dollars do not change. However, as indicated, beginning with the Dec-11 SAR, Base Year Dollars changed from 2002 to 2012.

It is appropriate to ensure that a firm dollar measuring rod is employed in the analysis over the SAR time periods employed, is to convert costs through Dec-11 SAR from 2002 Dollars to 2012 Dollars. One also needs to account for the fact that the APB changes on 26-Mar-12 are reported in the Dec-11 SAR, while the SAR Baseline changes to 26-Mar-12 are contained in the Dec-12 SAR. As shown in the table, Base Year Dollars shifts from 2002 to 2012 in the Dec-11 SAR.

Figure 8 compares RDT&E and Procurement Baseline Costs as the two baselines change. As indicated, Baseline costs remain constant throughout a baseline period. Also, of interest is the fact that over several APB periods, the Baseline costs are higher for the APB than for the SAR Baseline. One area affected by this occurrence relates to the Nunn-McCurdy cost breaches. Nunn-McCurdy compares cost estimates to APB costs. Given the curves shown in Figure 8, the DoD would have a bit more flexibility in avoiding a breach if SAR Baseline costs, in this situation, were employed as the objective baseline cost threshold.

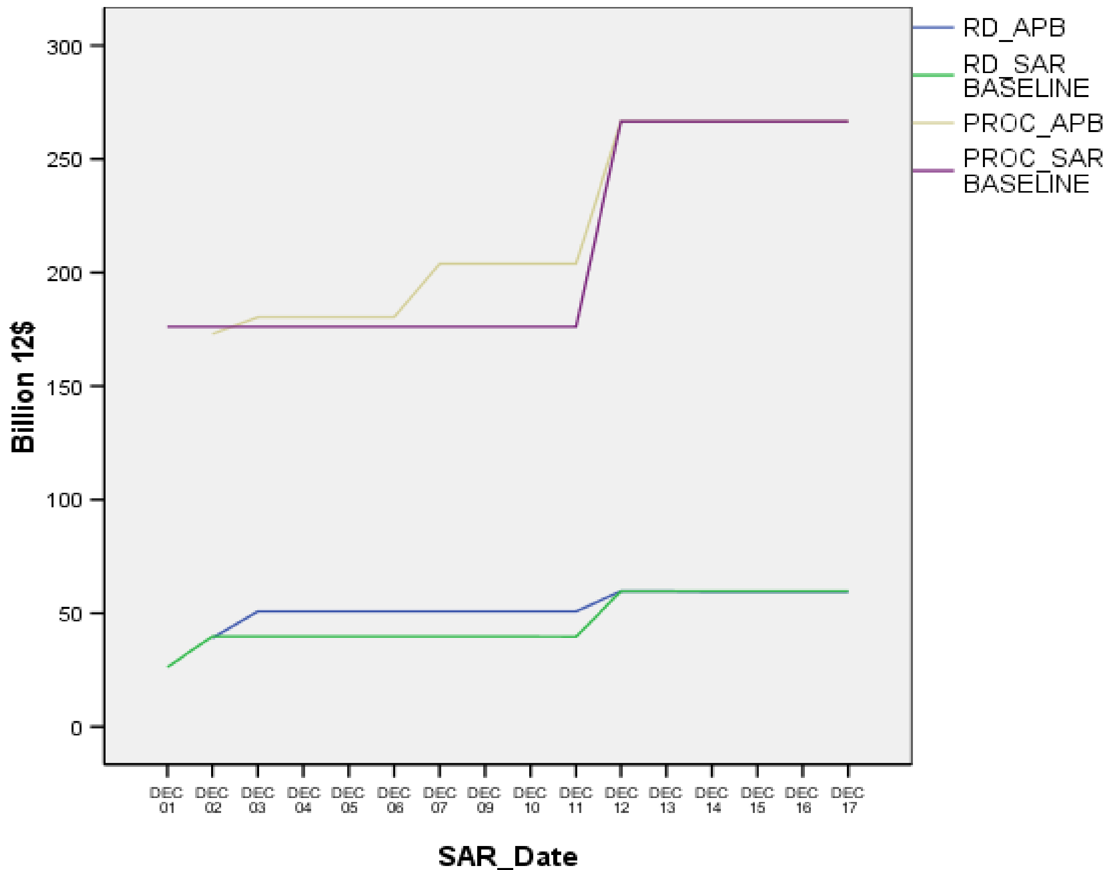

Figure 8. RDT&E and Procurement Baseline Cost Changes (2012 $B)

Next, the SAR date-specific Current Cost Estimates are shown in Figure 9 with some relatively mild fluctuations about a clear upward trend. As seen in the following charts, these current estimates are determined by starting with the SAR Baseline costs and augmenting these costs with the relevant Program Variances.

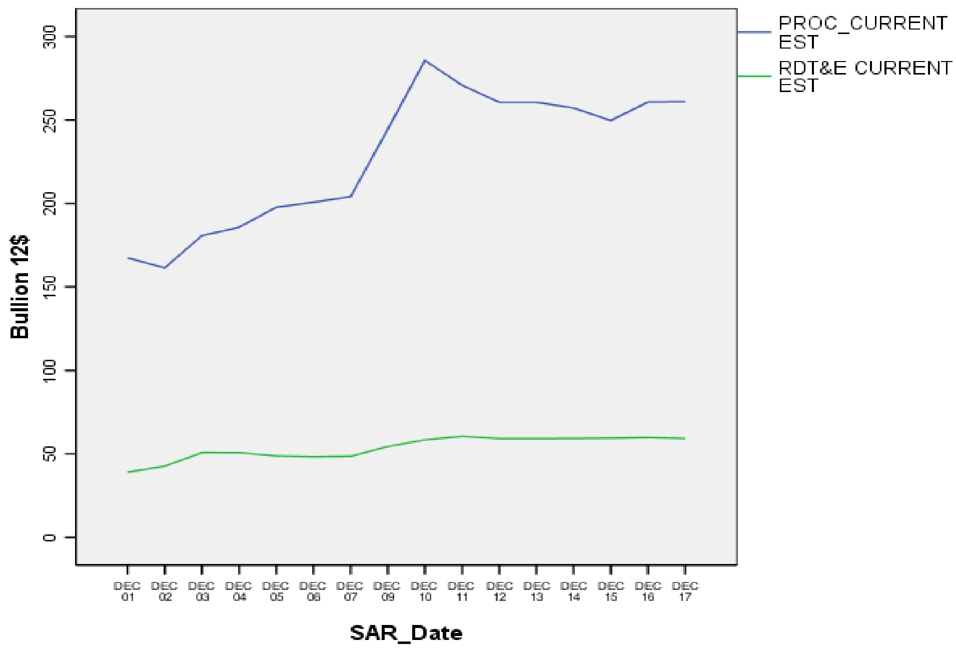

Figure 9. Current Cost Estimates for RDT&E and Procurement by SAR Date

Current variances are fairly volatile from to year, and this provides an explanation as to why there is some variation in Procurement and RDT&E costs along an identified upward trend. Only, for procurement from SAR-2007 through 2009 was there a large increase estimated.[iii] However, estimated procurement costs decline fairly quickly and return to historical trend. The variation about the trend may even suggest that the estimation of current procurement cost is based on the variances rather than estimated program being first estimated, and variance levels determined that are consistent with current estimated costs.

Program Variance

Several categories of Program Variance have been defined. Given that this analysis addresses schedule increases for the F-35, it would seem that Schedule Variance would provide one of the keys for understanding increases in schedule milestones. However, increases in Engineering Variance can impact schedule if the “physical or functional characteristics” become more difficult to achieve. Quantity Variance can also impact schedule. if there is an increase or decrease in the quantity produced.

Table 5 contains the categories and associated definitions. For each of these categories, both prior variance and current variance are identified. One should also note that when the variance data are converted to constant dollars, then Economic Variance, which accounts for changes in price level, is not an appropriate category in constant dollar analysis.

Table 5. SAR Program Variance Categories and Descriptions

SAR Program Variance ($)	Descriptron
Quantity	Change in mumber of units acqired
Schedule	Change in procurement or delivery schedule, completion date,
	or iintermediare milestones for development or production
Engineering	Change in the physical or functional characteristics of a system
	or item delivered
Estimating	Change due tro corrections in previious estimating errors or
	refnements of current estimate
Orher	Change due to unforseen events or not covered in any other
	category (e,g,, national disaster or strrike)
Support	Change associated with the support requirements for major item
	of hardware
Economic	Change in price level
Source: Based on definitions provided by Paul Hough in Pitfalls in Calculating Cost Growth from	
Selected Acqjisn Reports, The RAND Corp., N-3136-AF, 1992.	

An important point is that current variances represent changes in cost over the remainder of the program. Furthermore, the appropriation type of the variances (RDT&E versus Procurement) does not necessarily respond to the appropriation associated with a program activity within which these variances are determined.

SAR Baseline costs represent the “baseline” cost level that initiates a current estimate, and these costs remain unchanged until the next SAR Baseline is established. Only two SAR Baseline Dates are indicated for the F-35: October 26, 2001, and March 26, 2012. Recognizing this, we now discuss how program variance information is used to develop current cost estimates for both RDT&E and Procurement. This is best described with series of steps:

1. Obtain Prior Program Variance information since the start of the current SAR Baseline by Variance category for both RDT&E and Procurement. As will be discussed, this equals the Total Variance information reported in the previous year's SAR.
2. Aggregate these prior variances across all categories to obtain subtotals for Prior Variance RDT&E and Procurement.
3. Determine Current Variance levels by variance category for both RDT&E and Procurement that are applicable to the current SAR period and compute a subtotal for Current Variance costs.
4. Add the RDT&E and Procurement subtotals for Prior Variances and Current Variances to obtain "Total Variance" by appropriation category. These totals represent the total change in cost generated by the Prior and Current Variances.
5. Add the resulting total changes in RDT&E and Procurement Variances to the SAR Baseline Objective Costs for RDT&E and Procurement.
6. The resulting totals equals the Current cost estimates for RDT&E and Procurement.

This is how the data reported in Figure 8 are obtained. It is important to reiterate that subsequent SARs within the SAR Baseline period do not begin with the previous year's Current Cost Estimate. Within a SAR Baseline period, costs always begin with the SAR Baseline Costs that do not change during a particular SAR Baseline.

There is an important relationship that is implicit in the identified procedure:

$(\text{Total Variance})_{t\ t} = (\text{Prior Variance})_{t-}$

Because $(\text{Prior Variance})_t$ equals the lagged $(\text{Total Variance})_{t-1}$, and this Total Variance can be calculated using data from the previous SAR, $(\text{Prior Variance})_t$ can be viewed as a lagged variable. As shown below in the empirical section of the analysis, when $(\text{Prior Variance})_t$ is included in a model as an explanatory variable, this variable can also be viewed as lagged independent variable, and, as a result, aid in forecasting the dependent variable.[iv]

In the bar graph displayed in Figure 10, Current Cost Variances for RDT&E are aggregated across all SAR Years employed in the analysis. When interpreting the chart however, it is important to understand that RDT&E costs almost exclusively occur through the Dec-11 SAR. Therefore, the variances in Figure 10 only impact cost estimates during the Dec-01 through Dec-11 time periods.

One should not be surprised that RDT&E Quantity Variance is relatively low and that there is no RDT&E Support Variance for the F-35. Procurement is the Appropriation category that is primarily impacted by these two RDT&E variances.

RDT&E Schedule Variance, in contrast, would be impacted through changes in intermediate milestones and the effect of an early reduction in production quantity. RDT&E Engineering Variance can be expected to be large as a result of changes in performance characteristics, which ultimately may contribute to changes in the APB and SAR Baselines.

RDT&E Estimating Variance is quite large. This reflects the difficulties in estimating RDT&E costs for an advanced tactical fighter like the F-35. The interesting finding concerning Estimating Variance, however, applies to Current Procurement Variance and will be discussed shortly.

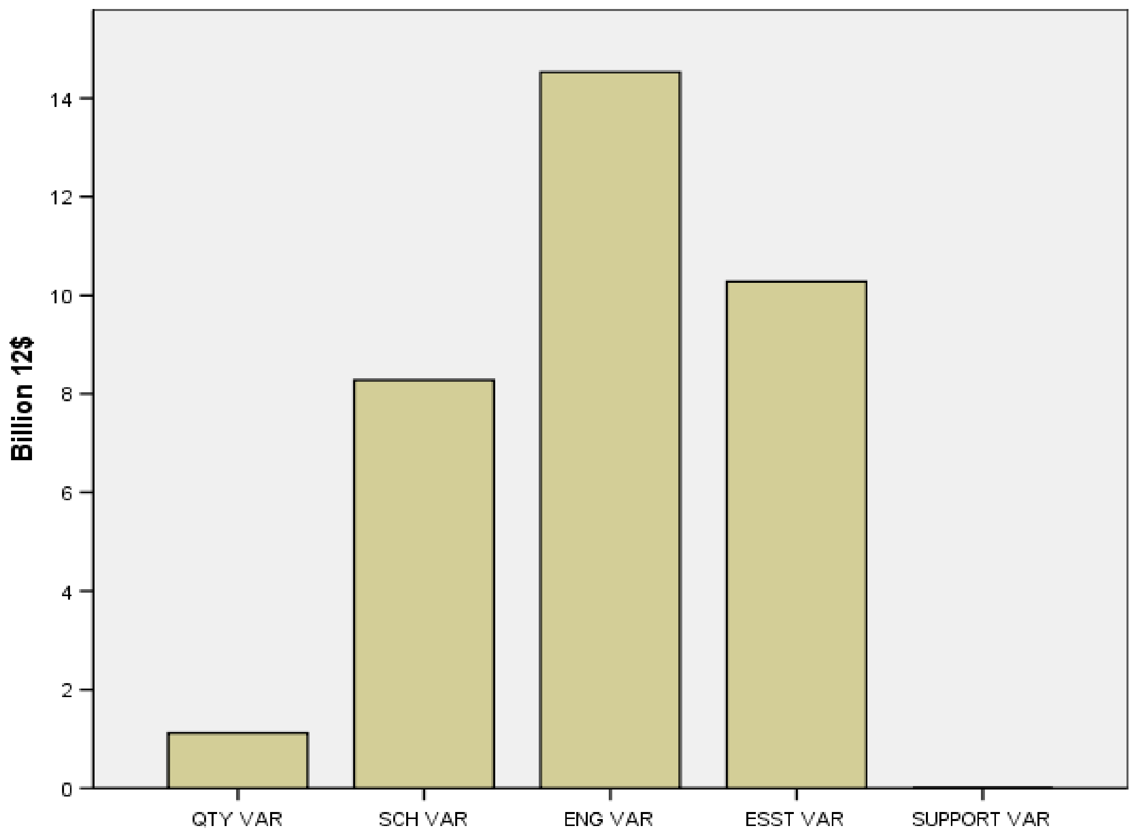

Figure 10. Aggregation of Current RDT&E Cost Variances for F-35

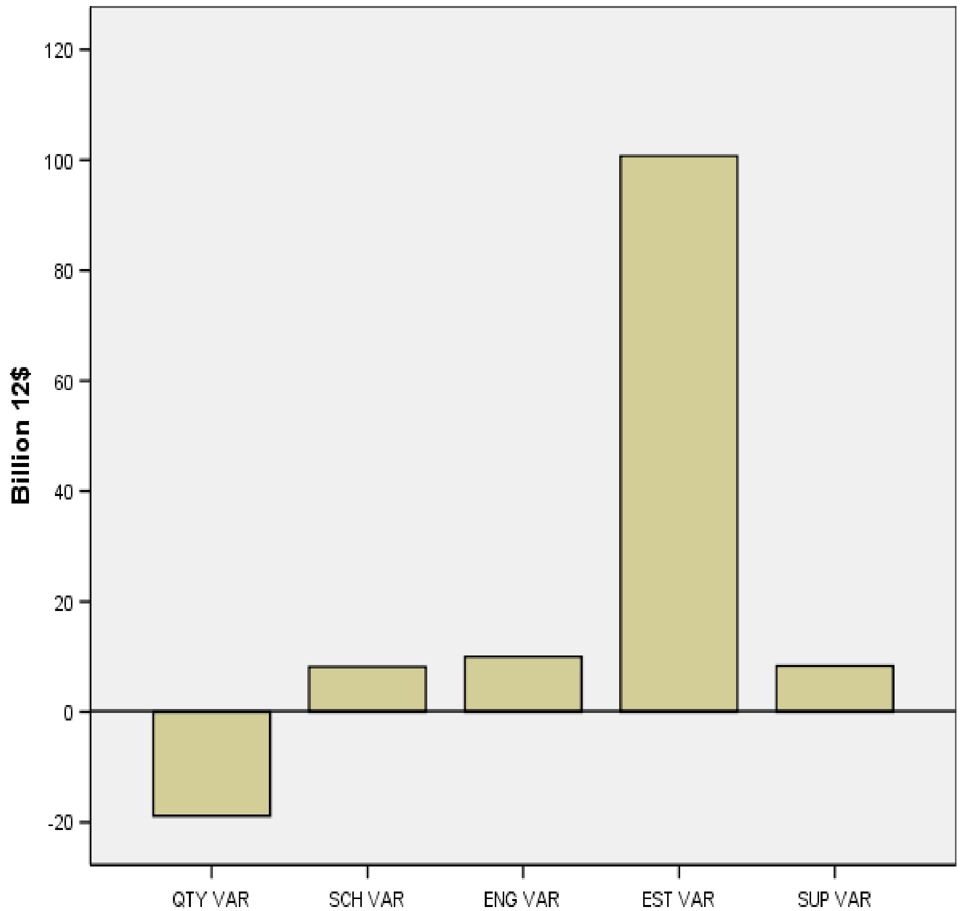

Figure 11. Aggregation of Current Procurement Cost Variances for F-35

When one turns to the Procurement Variances, the Quantity Variance can be explained by the reduction in total procurement quantities fairly early in the program. Refinements in one's understanding of support requirements can be expected to occur after most of the RDT&E activities have been completed. Perhaps the costs associated with support equipment may not have gone through the same rigorous planning process as the Air System and Propulsion prior to Milestone B, and also during the Engineering and Manufacturing Development (EMD) period.

What is extremely interesting is the high level of Estimating Variance. We should state again that Current Variance computations do not simply impact the current SAR year. The computations capture the effects of changes during a particular SAR year on costs incurred over the remaining life of the program. These costs would primarily be procurement costs, which suggests that there are difficulties in estimating these downstream costs.

Because of the large dollar value associated with aggregate procurement Estimating Variance, further analysis is appropriate.

Figure 12 compares the change in current procurement cost estimates with the aggregation of Current Procurement Estimating Variance during the first and second SAR Baseline Periods. As shown in Table 4, the first SAR Baseline Period begins October 26, 2001, and the second begins March 26, 2012. Therefore, Dec-01 SAR though the Dec-11 SAR report information from the first baseline period; the Dec-12 SAR through the Dec-17 SAR report information associated with the second SAR Baseline period.

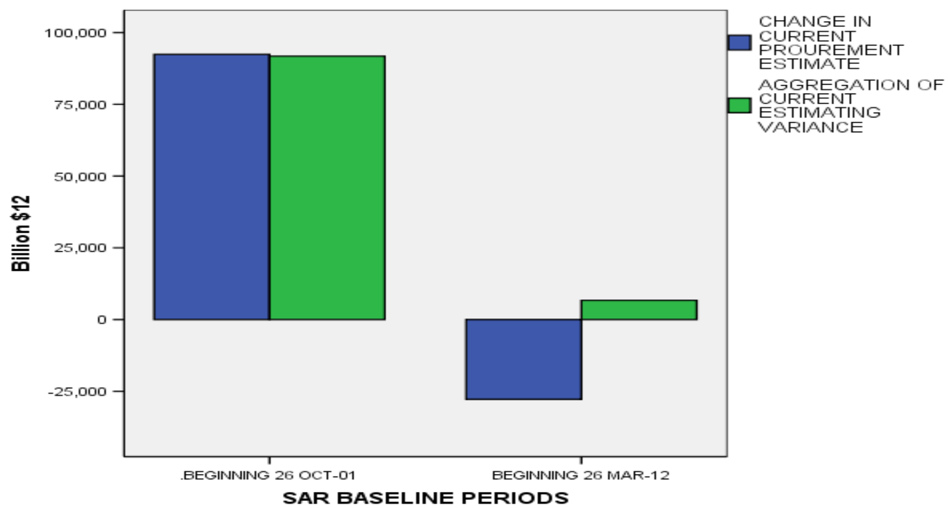

Figure 12. Comparison of Aggregate Procurement and Aggregate Estimating Variance During Two SAR Baseline Periods

Understanding the large size of Procurement Estimating Variance can be analyzed further by comparing the aggregate Estimating Variance cost with the increase in estimated procurement costs. The second SAR Baseline period begins March 26, 2012, and continues through the Dec-17 SAR. Therefore, the data in the first two bars are contained in SAR-01 through SAR-11; the second two bars on the right side of the diagram are based on data from SAR-12 through SAR-17, the second SAR Baseline period. During the first SAR Baseline period, which begins on October 2, 2001, all costs from SAR-01 through SAR-10 are RDT&E contracts with costs primarily associated with System Design and Development (SDD) contracts for the Air System and for Propulsion. In SAR-11, SDD RDT&E continues to have significantly higher RDT&E SDD costs for the Air System and Propulsion than Procurement costs. In this SAR, the first procurement contracts are identified for LRIP-2 through LRIP-4 in SAR-11.

Therefore, most of the Procurement Estimating Variances in the first SAR Baseline period are generated during the RDT&E phase of the program. These Estimating Variances represent growth in procurement costs for the duration of the program and indicate the difficulties in estimating procurement costs when Milestone B is approved, which is shortly before the start of the first SAR Baseline. These estimating variances may also represent information accumulated during the Manufacturing Development component of Engineering and Manufacturing Development (EMD) component during the first SAR Baseline.

We can also see in Figure 12 that the Aggregation of Procurement Estimating Variances across both SAR Baseline periods is larger than the change in Procurement costs. Figure 11 shows that Procurement Estimating Variance is approximately equal to the sum of the other four variance categories: Quantity, Schedule, Engineering, and Support.

So, what is the source of the changes in estimated procurement costs, which are lower than the Aggregate Procurement Estimating Variance? It seems that the Aggregate Estimating Variance is inconsistent with the change in Baseline Procurement Costs, when the Baseline changed from 26 Oct-01 to 26 Mar-12. The Aggregate Estimating Variance could have been too large, or the SAR Baseline procurement change could have been too small, or some combination of these two factors.

Clearly, however, the large dollar value of the Estimating Variance during the first SAR Baseline indicates that, for technologically advanced systems like the F-35, with a large number of procurement units whose production extends over numerous years, efforts to improve cost-estimation methodology should be pursued.

Program-Level Estimating Model

Table 6. Empirical Model Explaining MS C – MS B (2012$)

Explanatory Variables	Coefficients	t-stat
(Constant)	50.307	1.33
RD Prior Tot Variance ($M)	2.087	5.25
Proc Prior Tot Variance ($M)	0.160	4.09
SAR Baseline_D (SAR-12_SAR-17 = 1)	97.368	16.96
Appropriation _D (RDT&E = 1)	-10.022	-3.59
Total Procurement Quantity (DoD)	0.027	1.96
Dependent Variable: Milestone C - Milestone B (Months)		
R^2= .970; N=71		

To explain changes in the number of months between MS B and MS C, Prior Variance for RDT&E and Procurement are key explanatory variables and have the best statistical properties among the variance measures. And, as mentioned before, because this variable is really a lagged variable equal to Total Variance the prior period, the Prior Variance variables have attractive properties in forecasting the dependent variable. Furthermore, as shown below, when a program-level model is linked to a contract-level model, Procurement Prior Total Variance has the the best statistical properties for use as a variable linking contract-level data and program-level data.

Recall that the date in which the current SAR Baseline begins constitutes the boundary of a particular Prior Variance computation. The first SAR Baseline begins on 26 Oct-01, and the second SAR Baseline begins on 26 Mar-12. Prior Procurement computations for first Baseline would then be reported in SARs Dec-01 through Dec-11; in the second baseline period, reported in SARs Dec-12 through Dec-17, computations within the second SAR Baseline would be developed in SAR-12 through SAR-17.

This computational procedure is also related to SAR Baseline_D, a categorical variable which takes on a value of 1 during the second SAR baseline period, which begins March 26, 2012, and for which initial data are captured in SAR-12. This second SAR Baseline period continues through SAR-17. This variable shows that, other things equal, there was a significant increase in MS C – MS B during the second SAR Baseline period when MS C – MS B, which is affected by Prior Variances, is large.

The variable Appropriation-D takes on a value of 1 for RDT&E and 0 for Procurement. The negative coefficient for this variable is consistent with the fact that RDT&E expenditures occur primarily fairly early in the program when MS C – MS B is small. The larger changes in this difference occur during the period when Procurement expenditures are high.

Finally, the model shows that Procurement Quantity, which declines early in the program, and then increases around SAR 2016 is associated with a positive increase in MS C – MS B. The decrease in total procurement quantity is reported in SAR-03, and there is a significant negative procurement Current Quantity Variance reported in SAR-02. Furthermore, an increase in procurement quantities is reported in SAR-16 with an associated positive SAR-16 Current Quantity Variance. These changes in cost ultimately impact MS C – MS B and explain the positive regression coefficient.

Contract Analysis

We turn now to an analysis of contract data. Data are reported for 71 distinct contract observations associated with 28 different contracts identified in the Dec-01 through Dec-17 SAR reports. The difference arises because 14 of the contracts are reported as multi-SAR year contracts. For the 71 contract observations, the data associated with each of the multi-year contracts differ with respect to Contract Schedule, Contract Cost Variance, and other contract specific variables. Variance data are reported as Prior Contract Variance, Cumulative Variance to date, and Current Variance. Other data reported at the contract level are Award Date, Definitization Date, Contract Type, Initial Target Price, Current Target Price, Estimated Price at Completion, Initial Quantity, and Current Quantity. During the second SAR Baseline period, through SAR-17, procurement costs are high, and the data reported are for LRIP-2 through LRIP-10 for both the Air System and Propulsion subsystems.

We use only contract data provided in the annual SAR reports. First, we provide definitions of the Contract Variance variables. Then a graphical picture of the definitions in a certain context is provided. The graph should be an aid in understanding contract variance.

For a particular contract, Contract Schedule Variance (SV) is equal to the Budgeted Cost of Work Performed (BCWP) less the Budgeted Cost of Work Scheduled:

$$SV = BCWP - BCWS.$$

Contract Cost Variance equals the Budgeted Cost of Work Performed (BCWP) less the Actual Cost of Work Performed:

$$CV = BCWP - ACWP.$$

To simplify the relationship between contract and program variances, the following is a useful memory aid: with respect to cost, positive SVs and CVs are typically positive contract outcomes, whereas increases in Program Current Variances, which represent increases in estimated program cost, are typically negative outcomes.

In the SAR reports, BCWP, BCWS, and ACWP are not reported. In addition to SV and CV, Prior SV (PSV), Prior CV (PCV) Cumulative SV, and Cumulative CV though he current SAR year are reported:

$$SV = CSV - PSV, \text{ and}$$

$$CV = CCV - PCV$$

Once again, there is an observable time linkage between successive periods. The following must hold:

$$PSV_t = CSV_{t-1}, \text{ and}$$

$$PCV_t = CCV_{t-1}.$$

As before, these linkages aid in forecasting when an empirical contains either PSV or PCV as independent variables. While it is the case that individual contracts are operational during a shorter period of time than programs, linked data remain relevant. Nevertheless, the Contract Award Date, Contract Definitization Date, and SAR Year should be considered when contract data are being analyzed.

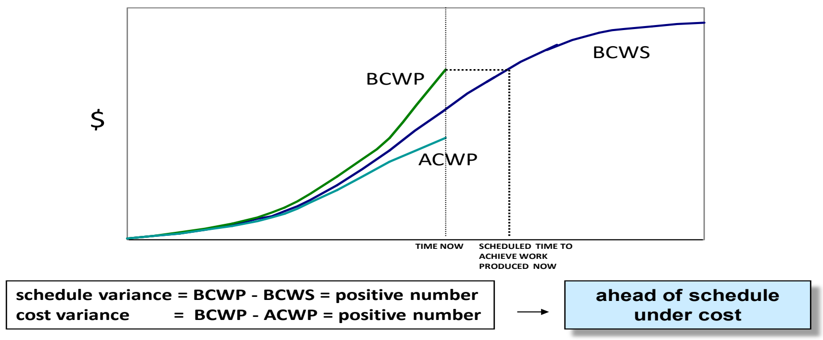

Figure 13. Graphical Depiction of Positive Schedule and Cost Variance
(Based on Eleanor Haupt briefing, "BAS105 Basic Earned Value Management.")

Associated with work scheduled during a contract is a budget developed by the contractor. This is shown as the Budget Cost of Work Scheduled (BCWS) curve.

For a particular contract, a certain amount of work has been performed, typically measured using milestones that have been met. These are represented as the Budget Cost of Work Performed (BCWP) curve. One can compare the budget actually expended with the amount of work scheduled to be performed for that budget level. If, during the current period, the budget associated with the work performed is greater than the budget associated with the work scheduled to be performed, the contractor is ahead of schedule. More directly, from the chart, one can see that if BCWP is greater than BCWS, the contractor is ahead of schedule.

One can compute Cost Variance by comparing BCWP and the Actual Cost of Work Performed (ACWP). If ACWP is less than BCWP, the contractor is under cost.

SAR reports do not contain BCWP, BCWS, or ACWP. However, SV and CV are reported as are Prior Cumulative Variances to Date minus Prior Contract Variance for both schedule and cost.

Empirical Contract Model

Table 7 summarizes the empirical results obtained with Contract and Program Data.

Table 7. Contract Model With Program Linkage

Explanatory Variables	Coefficients	t_Stat
(Constant)	7336.034	0.74
SAR Date-Contract Award Date (Days)	9.933	1.82
Prior Contract Cost Variance	-72.361	-1.65
Prior Contract Schedlule Variance	120.105	2.09
Dependent Variable: Proc Prior Variance (2012 $B)		
R^2 = .183, N = 55		

In developing a model containing contract data, one key objective is to link the contract-level data to a program variance indicator. This permits the ultimate computation of the effect of changes in certain contract data on MS C – MS B. We find that the program measure with the best statistical properties is the F-35 program's Prior Procurement Variance, which as we have discussed is equal to the one-year-lagged Total Procurement Variance.

An empirical analysis that achieves such a linkage is more difficult than working exclusively at the program level. In large part, this results from analyzing 71 contracts that have more variation than the program data. Therefore, the R^2 is significantly lower than is obtained in the previous program-level model. However, the t-statistics are acceptable for this type of regression.

We included as an explanatory variable length of time, in days, between contract award date and SAR date. We aren't certain when Initial Target Price is set. Based on experience, contractor work can begin prior to contract definitization by issuing a letter contract. Current Target Price would also occur at an uncertain time, but the SAR discussion suggests that this is the target price in effect during the SAR Year. As a result, SAR Date – Contract Award Date, in days, is selected as an explanatory variable. The longer this period, the greater the opportunity for changes, including target price, which could affect program-level Procurement Prior Variance.

The contract variance variables that most closely parallel Procurement Prior Variance are Prior Contract Cost Variance and Prior Contract Schedule Variance. As these contract variances equal zero at the start of the contract, the Prior Contract Variances begin aggregating at the start of the second contract year.

As indicated, the number of contract periods over which the prior variance calculations are performed is typically much lower than the length of a SAR Baseline period. Nevertheless, these contract variance measures, particularly Prior Contract Cost Variance, appear promising for explaining and predicting program-level Procurement Prior Variance.

When interpreting the regression coefficients linking Contract Variance and Program Variance, one needs to keep in mind the following: Positive changes in Program Variance categories reflect program cost increases, and as such are "bad," while positive increases in Contract Cost and Schedule Variances are "good" in the sense that they represent relative reductions in cost and improvements in schedule performance.

The negative coefficient for Prior Contract Cost Variance, therefore, would be consistent with expectations. If the Budget Cost of Work Performed increases faster than the Actual Cost of Work Performed, the contractor's costs relative to budgets would be decreasing, and this decrease should have a cost-reducing effect on Program-level Prior Procurement Total Variance.

However, Prior Schedule Variance has a positive coefficient indicating that when this variable decreases (a bad), so too does program-level Procurement Prior Variance (an ostensible good). To understand this anomaly, further research is required.

Observations and Recommendations

The following summarizes observations derived from this analysis. Some recommendations are also provided.

1. Analysis is a useful aid for understanding the meaning and uses of annual SAR data;
2. Data analysis and empirical models show the type of analysis that can be accomplished using exclusively annual SAR data;

3. There is evidence of significant difficulties estimating procurement costs at time of Milestone B approval;
4. Empirical model explaining months between Milestone B and Milestone C has attractive forecasting properties. The contract model also has reasonably attractive properties. However, the coefficient of Prior Contract Cost Variance in the contract model requires further analysis;
5. Future research would likely focus on policy response models that include variables whose adjustment can influence intra-program schedule;
6. Exploration of finer grain models that portray the causal structure associated with program schedule achievement is recommended; and
7. Systems Dynamics and other network models that include program schedules as an integral part of the modeled acquisition process should be investigated.

THIS PAGE LEFT INTENTIONALLY BLANK

References

Abdel-Hamid, T. K. (1990). Investigating the cost/schedule trade-off in software development. *IEEE Software*, *7*(1), 97–105. http://doi.org/10.1109/52.43055

Airpower Australia. (2016, March). Is the JSF really a fifth-generation fighter? Retrieved from http://www.ausairpower.net/jsf.html

Arnold, Scot et al. (2010), WSARA 2009: Joint Strike Fighter Root Cause Analysis, Washington, DC: Institute for Defense Analyses, Paper 4612, June.

Barrett, M., & Carpenter, M. (2017, July). *Survivability in the digital age: The imperative for stealth*. Arlington, VA: Mitchell Institute. Retrieved from http://docs.wixstatic.com/ugd/a2dd91_cd5494417b644d1fa7d7aacb9295324d.pdf

Blickstein, I., Boito, M., Drezner, J. A., Dryden, J., Horn, K., Kallimani, J. G., … Wong, C. (2011). *Root cause analyses of Nunn-McCurdy Breaches, Volume I*. Santa Monica, CA: RAND Corporation.

Bogdan, C. (2012, September 17). *Joint Strike Fighter requirements*. National Harbor, MD: AFA Air & Space Conference. Retrieved from http://www.af.mil/shared/media/document/AFD-120919-046.pdf

Brooks, F. P. (1995). *The mythical man-month: Essays on software engineering, Anniversary Edition, 2/E*. Boston, MA: Addison-Wesley.

Capaccio, A. (2018, March 28). Air Force risks losing third of F-35s if upkeep costs aren't cut. *Bloomberg News*. Retrieved from https://www.bloomberg.com/news/articles/2018-03-28/air-force-risks-losing-third-of-f-35s-if-upkeep-costs-aren-t-cut

Christensen, C. M., & Bartman, T. (2016). The hard truth about business model innovation. *MIT Sloan Management Review.* Retrieved from https://sloanreview.mit.edu/article/the-hard-truth-about-business-model-innovation/

Church, A. (2016, February 26). Sensor to trigger in one challenging step. *Air Force Magazine Daily Report*. Retrieved from http://www.airforcemag.com/DRArchive/Pages/2016/February%202016/February%2026%202016/Sensor-to-Trigger-in-One-Challenging-Step.aspx

Chairman of the Joint Chiefs of Staff (CJCS). (2015, January 23). *Joint Capabilities Integration and Development System (JCIDS)* (CJCSI 3170.01I). Retrieved from https://dap.dau.mil/policy/Documents/2015/CJCSI_3170_01I.pdf

Demeulemeester, E. L., & Herroelen, W. S. (2006). *Project scheduling*. Berlin, Germany: Springer Science & Business Media.

Deptula, D. (2016, September). *Evolving technologies and warfare in the 21st century: Introducing the "Combat Cloud"* (Mitchell Institute Policy Papers, Vol. 4). Retrieved from http://docs.wixstatic.com/ugd/a2dd91_73faf7274e9c4e4ca605004dc6628a88.pdf

Director, Operational Test & Evaluation (DOTE). (2018). FY17 programs, F-35 Joint Strike Fighter. Retrieved from http://www.dote.osd.mil/pub/reports/FY2017/pdf/dod/2017f35jsf.pdf

Franck, R., Hildebrandt, G., & Udis, B. (2016). *Toward realistic acquisition schedule estimates*. Presented at the Seventh Annual Acquisition Research Symposium, Monterey, CA.

Franck, R., Hildebrandt, G., Pickar, C., & Udis, B. (2017). *Realistic acquisition schedule estimates: A follow-on inquiry* (pp. 1–20). Presented at the 14th Annual Acquisition Research Symposium, Monterey, CA.

Franck, R., & Udis, B. (2016, August 23). Why America shouldn't build sixth-generation manned fighters. *National Interest*. Retrieved from http://nationalinterest.org/blog/the-buzz/why-america-shouldnt-build-sixth-generation-manned-fighters-1744

Friedberg, Sydney J.,Jr. (2018), Navy, Marine F-18s in 'Death Spiral' As Readiness Plummets, Breaking Defense, https://breakingdefense.com/2017/02/navy-marine-f-18s-in-death-spiral-as-readiness-plummets/, 8 February.

Global Security (2018), *DARPA / Navy Common Affordable Lightweight Fighter (CALF) 1993-1994*, accessed 22 May, https://www.globalsecurity.org/military/systems/aircraft/calf.htm

Government Accountability Office (GAO). (2017, October). *DoD needs to address challenges affecting readiness and cost transparency* (GAO 18-75). Retrieved from https://www.gao.gov/assets/690/687982.pdf

Gouré, D. (2018, March 15). Getting the Navy's next generation network right. *Real Clear Defense*. Retrieved from https://www.realcleardefense.com/articles/2018/03/15/getting_the_navys_next_generation_network_right_113197.html

Godlewski, E., Lee, G., & Cooper, K. (2012). System dynamics transforms Fluor project and change management. *Interfaces*, *42*(1), 17–32. Retrieved from http://doi.org/10.1287/inte.1110.0595

Harrigian, J., & Marosko, M. (2016, July). *Fifth generation air combat: Maintaining the joint force advantage* (Mitchell Forum No. 6). Washington, DC: Mitchell Institute. Retrieved from http://docs.wixstatic.com/ugd/a2dd91_bd906e69631146079c4d082d0eda1d68.pdf

Hendrickson, C., Martinelli, D., & Rehak, D. (1987). Hierarchical rule-based activity duration estimation. *Journal of Construction Engineering and Management*, *113*(2), 288–301. Retrieved from http://doi.org/10.1061/(ASCE)0733-9364(1987)113:2(288)

Herroelen, W. (2005). Project scheduling—Theory and practice. *Production and Operations Management*, *14*(4), 413–432.

Howick, S. (2003). Using system dynamics to analyse disruption and delay in complex projects for litigation: Can the modelling purposes be met? *Journal of the Operational Research Society*, *54*(3), 222–229. Retrieved from http://doi.org/10.1057/palgrave.jors.2601502

Hughes, R. T. (1996). Expert judgement as an estimating method. *Information and Software Technology*, *38*(2), 67–75. Retrieved from http://doi.org/10.1016/0950-5849(95)01045-9

Jaafari, A. (1984). Criticism of CPM for project planning analysis. *Journal of Construction Engineering and Management*, *110*(2), 222–233. Retrieved from http://doi.org/10.1061/(asce)0733-9364(1984)110:2(222)

Keijsper, G. (2007). *Joint Strike Fighter: Design and development of the international fighter*. Barnsley, England: Pen & Sword.

Kreps, D. (1990). *A course in microeconomic theory*. Princeton, NJ: Princeton University Press.

Lorell, Mark et al. (2013), Do Joint Fighers Save Money? Santa Monica, CA: RAND.

Morris, P., & Hough, G. H. (1988). *The anatomy of major projects*. Hoboken, NJ: John Wiley & Sons.

PERT High Tech Projects. (2005). PERT High Tech Projects, 1–6.

Project Management Institute (PMI). (2017). *A guide to the Project Management Body of Knowledge (PMBOK® Guide)—Sixth edition and Agile practice guide (ENGLISH)*. Newtown Square, PA: Project Management Institute.

Schwindt, C., & Zimmerman, J. (2015). *Handbook on project management and scheduling, Vol. 1*. Berlin, Germany: Springer.

Steidle, C. E. (1997, January–March). The Joint Strike Fighter program. *Johns Hopkins APL Technical Digest*, *18*(1), 6–20. Retrieved from http://www.jhuapl.edu/techdigest/TD/td1801/steidle.pdf

Sweetman, Bill (2012), Generation Gap, Aviation Week, 26 November, pp. 22-23.

Taylor, J. G. (1983). *Lanchester models of warfare, Vol II*. Arlington, VA: Operations Research Society of America.

Timperlake, E. (2017, September 3). Shaping a way ahead to prepare for 21st century conflicts: Payload-utility capabilities and the kill web. Second Line of Defense. Retrieved from http://www.sldinfo.com/shaping-a-way-ahead-to-prepare-for-21st-century-conflicts-payload-utility-capabilities-and-the-kill-web/

Tirpak, John (2011), New Life for Old Fighters, Air Force Magazine 94(2), February, http://www.airforce-magazine.com/MagazineArchive/Pages/2011/February%202011/0211fighters.aspx.

U.S. Air Force Enterprise Capability Collaboration Team. (2016, May). *Air superiority: 2030 flight plan*. Retrieved from http://www.airforcemag.com/DocumentFile/Documents/2016/Air%20Superiority%202030%20Flight%20Plan.pdf

Government Accountability Office (GAO). (2015). *Acquisition reform: DoD should streamline its decision-making process for weapon systems to reduce inefficiencies* (GAO-15-192). Washington, DC: Author.

Williams, T., Eden, C., Ackermann, F., & Tait, A. (1994). The vicious circles of parallelism. *International Journal of Project Management*, *13*(3), 151–155.

Williamson, Oliver E. (1996). The Mechanisms of Governance. New York: Oxford University Press.

---

[i] The graph is contained in *Toward Realistic Schedule Estimates*, Raymond E. Franck, Jr., et al., presented at the 13th Annual Acquisition Research Symposium, Monterey, CA, NPS, April 30, 2016.

[ii] Should Schedule is discussed in a presentation given by then-secretary of the Air Force on September 14, 2014, to the Air Force Association's Air and Space Conference.

[iii] The one year for which a SAR report is not available is Dec-08. There was an extensive search, and comments on the Internet suggest that this report is difficult to obtain (if it in fact exists). The straight line indicating procurement cost growth from SAR-07 to SAR-09 is an artifact of the missing data. Using the logic of program and schedule variance and the nature of Baselines, much of the missing data can be computed, and these additions will be incorporated into the next iteration of this paper.

[iv] When the SAR Baseline changes, a new Baseline Cost Estimate is determined, and Prior Variances are set to zero. This effect may be fully visible only at the quarterly SAR level. However, the same principle that connects Prior Variance in one period with the Total Variance in the previous period continues to apply. The Current Variance, during the first year of new Baseline period, becomes the Prior Variance for the second year of the new Baseline period.