Schedule and Cost Estimations Through the Decades: Are They Improving?

Captain Sammantha Jones

Edward D. White, Ph.D.

Jonathan D. Ritschel, Ph.D.

Shawn M. Valentine

Acquisition professionals strive to provide the best estimation of schedule and cost to deliver war-winning capability. Numerous reforms and improvement initiatives have been implemented towards improving these cost and schedule outcomes in Major Defense Acquisition Programs (MDAPs). This leads to the following question: Are schedule and cost outcomes improving over time? We use both descriptive and inferential techniques to investigate schedule and cost trends in MDAPs from the 1970s to 2010s. We find schedule growth does not exhibit statistically significant improvement across the decades; all decades indicated a consistent schedule slippage for a typical MDAP. In contrast, the analysis of Cost Growth Factors (CGFs) detected statistical differences in some instances. The most novel finding, however, is found in the *standard deviations* of CGFs. We identified a statistically significant decreasing trend in the standard deviations of total program CGFs throughout the decades. This lowering variability trend also appeared for Program Acquisition Unit Cost (PAUC) CGFs from the 1980s onward. The decrease in variability of cost estimates suggest to us that cost estimators and/or the process behind them might be improving over time.

This article identifies macro-level trends of cost and schedule growth for Major Defense Acquisition Programs (MDAPs) from the 1970s to the 2010s. Specifically, we investigate overall program cost growth, program acquisition unit cost (PAUC) growth, and schedule growth for the Department of the Defense's (DoD) largest program acquisitions. The inspiration for this study came from Arena et al. (2006) and Younossi et al. (2007). Both papers provide insights into cost growth of MDAPs mainly prior to 2000. This article can be considered an extension of these often-cited works with a few key differences.

We do not delineate between development and procurement costs; we consider these together as total program cost as reported in the Selected Acquisition Reports (SARs). Although dividing cost growth into development and procurement is a common practice when analyzing MDAPs, we wanted to look at the overall cost and schedule growth holistically. There are other deviations between the analyses, such as how the data is presented and the type of inferential analyses utilized, but the overall goal of this paper is to investigate cost and schedule growth from the 1970s to the 2010s and to determine statistically whether the DoD has seen a change in cost or schedule growth over this timespan.

Background

MDAPs are essential for the development and production of military aircraft, satellites, missiles, and other large investment items that U.S. military operations require. By statute, MDAPs are categorized as Acquisition Category I (ACAT, 2021) programs if they have either total expenditure of research, development, test and evaluation (RDT&E) costs greater than $525 million (fiscal year 2020 constant dollars), total expenditure of procurement costs greater than $3.065 billion (fiscal year 2020 constant dollars), or specifically designated by milestone decision authority as special interest (MDAP, 2020). MDAPs are the DoD's largest investments and constitute a large proportion of the DoD portfolio relative to their program numbers. These investments often entail large economic risks.

Currently, the Government Accountability Office (GAO) reports annually on DoD weapon systems based on their total cost and acquisition status. Of the 107 programs evaluated in their 2021 report, 84 were MDAPs. These 84 MDAPs have a total planned investment of 1.79 trillion Fiscal Year (FY) 2021 dollars. The GAO has reported consistent cost growth in the DoD's MDAP portfolio for the last 15 years. They attribute the most dramatic cost changes to quantity changes (Government Accountability Office, 2021). Other studies have also noted historical precedent for underestimating program costs (Arena et al., 2006; Younossi et al., 2007) and schedules (Monaco & White, 2005; Riposo, McKernan, & Kaihoi, 2014). Light et al. (2017) even recommended that the acquisition community approach early cost estimates with skepticism.

Cost growth in MDAPs appears common; however, dramatic growth within programs can lead to a Nunn-McCurdy Breach. From 1997 to 2016, 58 out of 189, or 36% of MDAPs experienced cost growth large enough to cause such a breach. Out of these 58 breaches, 18 were significant and 40 were critical (USD(AT&L), 2016, p. 65). Significant breaches occur when current cost estimates meet or exceed 15% of the current baseline estimate or 30% of the original baseline estimate of an acquisition program. Critical breaches occur at the 25% and 50% levels respectively (Nunn-McCurdy Breach, 2021).

MDAPs that experience Nunn-McCurdy breaches are extreme examples of cost growth. But due to their programmatic costs, even a small cost growth percentage can add millions of dollars worth of additional funding needs for the programs. Schedule growth in MDAPs can also lead to readiness issues and apprehension for military and congressional leadership. Because of these funding and readiness issues, there have been efforts over the last several decades to reduce cost and schedule growth within MDAPs (Fox et al, 2011). These efforts include sweeping reforms, changes in business practices, updates to record keeping requirements, and adjustments in the overall structure of how MDAPs are executed, and their records maintained (Fox et al., 2011, Dwyer et al., 2020).

Over the last few decades there have been extensive analyses on DoD MDAPs. Various organizations such as the Congressional Research Service, the DoD itself, GAO, or even contracted organizations such as RAND or the Institute for Defense Analyses (IDA) have conducted these studies. In 2016 the DoD published an annual acquisition system performance report. In this report they analyzed MDAPs through a variety of different lenses to include cost and schedule growth, cost performance overall, cost performance broken out by development and production, cost growth by military departments, cost growth by contractors, and a few other viewpoints (USD(AT&L), 2016).

The 2016 report claims there has been a continuing improvement in the field of defense acquisitions, however their analyses concentrate on several various micro-level insights into the cost and schedule growth of DoD MDAPs. While these micro-level assessments are extremely important to understanding what is happening in specific MDAPs, their study does not provide a macro-level analysis truly examining whether the overall cost and schedule growth of MDAPs have changed over time (USD AT&L, 2016). Thus determining changes, if any, to cost and schedule growth is the intent of this article.

Data and Methods

Data

We utilized the Cost Assessment Data Enterprise (CADE) system to obtain the data for this article's analyses. Available since February of 2019, the CADE Selected Acquisition Report (SAR) database is a consolidation of DAMIR (Defense Acquisition Management Information Retrieval) SAR data and non-DAMIR legacy SARs. Using the SAR Unit Cost Report along with the Current and Baseline Estimate report and the CADE SAR Data listing, we identified 409 potential programs to analyze as of October 2021. From there, we excluded programs. Note that the dataset only includes MDAPs. Major Automated Information Systems (MAIS) are not part of the analysis. Table 1 lists the reasons for program exclusion and rationale. For programs categorized as transitioned or restructured, if these actions led to the creation of a new MDAP, then that new program remained in the database. For example, the WIN-T, after being broken into three separate programs, drove the creation of one MDAP that met the requirements to be included into our final dataset: the WIN-T increment 2.

SAR Sample Inclusion & Exclusion Table	
Total Number of SARS Available in CADE	409
Programs Classified as Terminated	26
Transitioned or Restructured Programs	11
SAR not Classified as MDAPs*	17
SAR w/no data available in CADE**	25
SARs with Missing Milestone B Data***	129
Programs < 5 years since MS B	7
Final MDAP SAR Sample	**194**

Table 1. Selected Acquisition Report (SAR) inclusion and exclusion table.

**This includes Pre-MDAP, Other, Special Interest, MAIS Major System, and DoE Program Classifications*

*** These programs were listed in CADE but had no cost or schedule data available for analysis*

**** These programs did not have any MS B data available as a starting point for the cost and schedule growth analysis*

We use Milestone (MS) B as the starting point for collecting program data, as this is typically considered the official start of a program (AcqNotes, 2021). Additionally, many previously published studies have used MS B as the starting point of their analyses on MDAP cost or schedule variations. These include studies by Younossi et al. (2007), McNicol (2018), and Dwyer et al. (2020).

The final exclusion criteria for our analysis involved accounting for the low maturity level of modern MDAPs. Programs that were less than five years old (and had yet to meet Initial Operating Capability (IOC)) were omitted from the analysis. This is because of the increased likelihood of these less than mature programs not having yet realized their schedule and cost changes compared to programs further along in development/ production. Within our schedule database (described later) the mean time for a MDAP to move from MS B to IOC is 8.6 years with 97 of the 120 taking more than five years to reach IOC. This maturity requirement led to the exclusion of seven MDAPs that reached MS B in 2017 or later. Younossi et al. (2007) adopted a similar exclusion criterion. Table A.1 in the Appendix lists all the programs by decade we analyzed for this article.

After the completion of this initial 194 MDAP database, we parsed the data into three separate databases to explore schedule growth, total program cost growth (this is just RDT&E plus procurement costs), and PAUC cost growth individually. To calculate the change in MDAP schedule growth, we used two main milestones: MS B and IOC. From the starting 194 programs, 74 did not have IOC estimates available in CADE and were subsequently not used in the schedule analysis. This left us with 120 programs for comparing schedule growth across the decades.

For investigating the Cost Growth Factor (CGF) – we define this shortly – for overall program total, a program was required to have cost data at MS B as well as on the last reported SAR. Eleven programs were missing cost data, reducing the initial 194 to 183 MDAPs for analyzing the CGF with respect to total program growth. For analyzing changes in PAUC, a program also needed quantity data. This 183 was further reduced to 165 since 18 MDAPs were missing quantity data. Table A.1 highlights these three databases used for comparing schedule, total program, and PAUC growth over the decades.

Completed Vs. Ongoing Programs	
Schedule Difference (Yr) - Completed vs. Ongoing	
Completed Programs	70
Ongoing Programs	50
Total	**120**
Overall CGF - Completed vs. Ongoing	
Completed Programs	118
Ongoing Programs	65
Total	**183**
PAUC CGF - Completed vs. Ongoing	
Completed Programs	102
Ongoing Programs	63
Total	**165**

Table 2. Completed vs. ongoing program breakout by response.

Besides initially analyzing all MDAP data together (completed and ongoing), we also split the completed and ongoing programs into separate categories. We do this to compare any aggregate statistical trends detected. Table 2 highlights the number of programs with respect to completed and ongoing. We define completed as any MDAP that no longer reports any SAR information. Ongoing is just the opposite. Those ongoing MDAPs still report SAR data even for programs that might have had a MS B date decades ago. This is because of ongoing contracts still reporting on those MDAPs.

After finalizing our three databases, we standardized all the cost data. Since these MDAPs can take many years to complete, there are instances where their costs are re-baselined to a different Fiscal Year (FY). There were several programs that had their estimates at MS B set to an earlier FY, while the current estimates were in a different FY. To ensure internal consistency for a program, we used the current base years for that program and standardized all cost data to that particular year. We used the Secretary of the Air Force raw inflation indices to perform these calculations.

Responses

In our analyses, we compared how three responses have changed from 1970s to the 2010s. These three responses consist of changes in schedule, total program cost, and PAUC. Equation (1) defines the percent schedule growth utilized. The denominator reflects the time from MS B to the last reported IOC date, while the numerator reflects the time from the estimated IOC date provided at MS B to the last reported IOC date. A value of 0 indicates no schedule growth. A positive percentage highlights a schedule slippage, while a negative percentage indicates a program reaching IOC quicker than expected at MS B.

$$(\text{IOC date}_{\text{Last Reported}} - \text{IOC date}_{\text{Estimated at MS B}}) / (\text{IOC date}_{\text{Last Reported}} - \text{MS B date}_{\text{Actual}}) \quad \textbf{(1)}$$

$$\text{Total Program Cost}_{\text{Last Reported}} / \text{Total Program Cost}_{\text{Estimated at MS B}} \quad \textbf{(2)}$$

$$\text{Total \# of Units}_{\text{Estimated at MS B}} / \text{Cost Estimate}_{\text{Estimated at MS B}} \quad \textbf{(3)}$$

$$\text{Total \# of Units}_{\text{Last Reported}} / \text{Cost Estimate}_{\text{Last Reported}} \quad \textbf{(4)}$$

To analyze total program cost growth, we took the last reported total cost value and divided it by the estimated total program cost at MS B (or equivalent from acquisition programs from earlier time periods). Equation 2 displays this calculation that generated the CGFs for our analysis. A CGF of 1 equates to a program experiencing no change in total program cost from MS B to the latest SAR. A value less than 1 suggests the program costs less than estimated at MS B, while a value greater than 1 shows an increased total program growth. This CGF calculation has been utilized in previous cost growth studies (Arena et al., 2006; Younossi et al., 2007; Kozlak et al., 2017).

The last response analyzed focused on the unit level, specifically at the PAUC. Quantity changes could drive some cost growth within MDAPs. To analyze the PAUC changes, we divided the total number of units estimated on the MS B SAR by the total cost estimate on the same SAR. [Note: The total number of units includes development and production units.] Then we calculated the current PAUC by taking the quantity reported on the latest SAR and dividing that by the latest program cost. Equations 3 and 4 highlight these calculations. After those two values were determined, we then divided (4) by (3) to arrive at the PAUC CGF, similar to the logic of (2).

Statistical Analysis

The goal for our analysis is to compare the decades, 1970s to 2010s, with respect to schedule growth, total program CGF, and PAUC CGF. We conduct these analyses with all MDAPs (completed and ongoing), then with only completed programs, and finally just ongoing programs. These analyses consist of a combination of descriptive and inferential statistics. The descriptive statistics include reporting means, medians, standard deviations, coefficient of variations (CVs), and interquartile ranges (IQRs) by decade.

Regarding inferential analyses, the standard F-test conducted under an Analysis of Variance was originally thought to be the best methodology to compare the responses across the decades. However, the non-normality pattern of the data indicated a non-parametric approach would be more appropriate given such inferential techniques have no distributional assumptions. Consequently, we utilized the non-parametric Kruskal-Wallis (K-W) test to determine statistically significant differences in the responses across the decades (Laerd Statistics, 2018). The specific null hypothesis tested is that the responses across the decades are equivalent versus the alternative hypothesis that at least one decade performs differently than the others. If the null hypothesis is rejected, then we use the non-parametric Steel-Dwass (S-D) pairwise comparison to isolate the specific decade(s) that is/are different.

The K-W and S-D inferential non-parametric tests are concerned with the typical response of a variable of interest. To assess how the variability of our responses (schedule growth, total program CGF, and PAUC CGF) might change across the decades, we employed the Brown-Forsythe (B-F) test. The B-F tests whether the response standard deviations/variances are equal or different across the decades. The B-F analyzes deviations based on the medians rather than the means of the data to minimize the effect of outliers or skewness in

Completed Vs. Ongoing Programs						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	12	24%	24%	0.15	0.27	0.65
1980	21	21%	20%	0.20	0.36	0.93
1990	27	32%	21%	0.28	0.44	0.87
2000	35	32%	21%	0.50	0.31	1.58
2010	25	13%	8%	0.31	0.28	2.34

Table 3. Schedule growth percentage summary statistics – Ongoing and completed programs. Median and mean values converted to percentages

the data (Brown & Forsythe, 1974, Stephanie, 2015). Since our data is not normally distributed, utilizing the B-F test provides more robust results versus the Levene Test, which uses means in its calculation. A level of significance of 0.05 was the default value that we used for all inferential hypothesis tests.

Analysis and Results

Total

The first analysis entailed all data, combining completed and ongoing programs. Table 3 presents the descriptive statistics for schedule growth by decade. All the means and medians are positive indicating consistent schedule slippage throughout the 1970s to 2010s for the typical MDAP program. The K-W and B-F tests returned *p*-values of 0.2123 and 0.6198, respectively, indicating no statistical difference among the decades with respect to the typical amount of schedule growth.

Table 4 presents the total program CGF by decade with one MDAP removed from the 1980s. This program is the DDG 51, the Arleigh Burke-class guided missile destroyers. Originally this MDAP had an initial purchase quantity of 14 ships; however, the most recent SAR shows the program acquiring 95. Such a dramatic change in units is more indicative of a scope change and not an issue with development/production issues. Thus the data are insufficient to parse cost/schedule increases from the quantity increase.

The K-W and B-F tests returned p-values of 0.0812 and 0.0006, respectively. The p-value of 0.08, although not significant at the 0.05 level, does suggest that there may be evidence that the 1970s possessed higher total program CGFs than the other decades if one was willing to increase the level of a Type I error to 0.10. The very low p-value of 0.0006 for the B-F strongly suggests statistical differences among the standard deviations of total program CGF by decade. As seen in Figure 1, the standard deviation has been

Overall CGF Summary Statistics - Ongoing & Completed (Excluding DDG 51)						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	29	2.83	1.44	3.62	2.88	1.28
1980	45	1.54	0.98	2.11	1.24	1.37
1990	37	1.66	1.26	1.17	1.51	0.71
2000	42	1.33	1.12	0.71	0.53	0.54
2010	29	1.14	1.02	0.37	0.25	0.32

Table 4. Overall CGF summary statistics – Ongoing and completed programs (Excluding the DDG 51 MDAP).

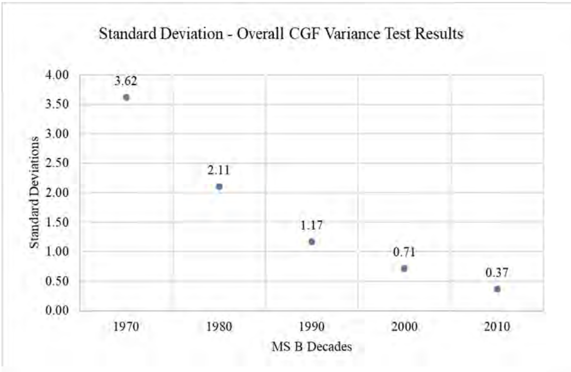

Figure 1. Standard deviations - Overall CGF of ongoing and completed programs.

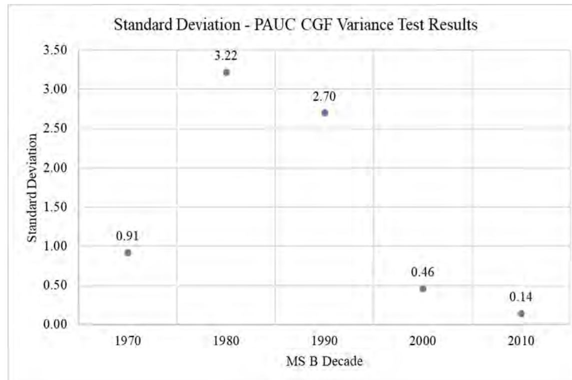

Figure 2. Standard deviations - PAUC CGF of ongoing and completed programs.

decreasing over the years. This appears to be a novel finding we haven't seen before in the literature. From our perspective, we have read many studies that have documented the patterns of cost and schedule growth. However, we have seen none that documented the actual variability of this process.

Table 5 presents summary statistics of the PAUC CGF by decade with one MDAP removed from the 2000s. This program is the C-130 AMP, which originally planned to acquire 519 units, but reported only purchasing nine (see Defense Industry Daily (2014) for some background on the decreasing number of units) on its most recent SAR. This drove PAUC from $7.26 million dollars per unit to $255.18 million dollars per unit. Since this outlier is markedly different from any other programs analyzed, we removed this program prior to conducting any inferential analysis.

The K-W and B-F tests returned p-values of 0.0302 and 0.0101, respectively. The K-W test concludes at least one decade is statistically different than the others with respect to PAUC CGF . The subsequent D-W test reveals that the 1970s and 1990s are statistically different than the 2010s with p-values of 0.0505 and 0.0411, respectively. These two decades have higher median PAUC CGFs than the lowest median PAUC CGF of the 2010s. With respect to the variability of PAUC CGF, Figure 2 highlights the standard deviations by decade; specifically, the 1980s and 1990s have statistically higher standard deviations than the other three decades while the 2000s and 2010s are decreasing.

Completed

We now duplicate the prior analysis but restrict it to just completed MDAPs. Table 6 presents the descriptive statistics for schedule growth by decade prior to any exclusions. The 2010s have only two MDAPs, indicating too few data to draw

PAUC CGF Summary Statistics - Ongoing & Completed (Excluding C-130 AMP)						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	25	1.55	1.36	0.91	1.10	0.59
1980	35	2.39	1.10	3.22	0.94	1.35
1990	37	2.15	1.26	2.70	0.99	1.25
2000	38	1.17	1.11	0.46	0.38	0.39
2010	29	1.01	1.02	0.14	0.19	0.14

Table 5. PAUC CGF summary statistics – Ongoing and completed programs (excluding C-130 AMP).

Schedule Growth Percentage Summary Statistics - Completed MDAPs						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	12	24%	24%	0.15	0.27	0.65
1980	20	20%	20%	0.19	0.34	0.97
1990	23	27%	18%	0.26	0.23	0.99
2000	13	34%	11%	0.66	0.38	1.95
2010	2	4%	4%	0.06	0.08	1.41

Table 6. Schedule growth percentage summary statistics – Completed programs. Means and medians converted to percentages.

Schedule Growth Percentage Summary Statistics - Completed MDAPs (Excluding Joint MRAP)						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	12	24%	24%	0.15	0.27	0.65
1980	20	20%	20%	0.19	0.34	0.97
1990	23	27%	18%	0.26	0.23	0.99
2000	12	17%	11%	0.22	0.30	1.35

Table 7. Schedule growth percentage summary statistics – Completed programs (excluding Joint MRAP). Means and medians converted to percentages.

any meaningful conclusions about this decade. In addition, there is a noticeable outlier in the 2000s belonging to the Joint Mine Resistant Ambush Protection (MRAP) MDAP; its schedule growth was approximately 240%. Table 7 presents the remaining descriptive data after removing these three programs and remain excluded for the K-W and B-F tests. The K-W and B-F tests returned p-values of 0.5208 and 0.3340, respectively, indicating no statistical difference among the decades with respect to the amount of schedule growth. This conclusion is consistent with using both completed and on-going MDAPs.

Table 8 presents the total program CGF by decade with the 2010s again removed (just two MDAPs completed). The K-W and B-F returned p-values of 0.1302 and 0.0270, respectively. The p-value of 0.1302 suggests that the decades are similar with respect to total program CGF, but the 0.0270 for the B-F suggests that the variability is not equal. As seen in Figure 3, there appears to be a decreasing trend in total program CGF variability by decade; a trend we witnessed in Figure 1.

Table 9 presents summary statistics of the PAUC CGF by decade with again the 2010 MDAPs (only two) removed and the exclusion of the C-130 AMP MDAP from the 2000 decade. The K-W and B-F tests returned p-values of 0.3508 and 0.4275, respectively. These results suggest no statistical differences with respect to the PAUC CGF (values or standard deviations) for the 1970s to the 2000s. This result is contradictory to what we concluded with all the programs, completed and ongoing. This suggests that the next section may reveal that PAUC CGF mainly varies between the decades for just ongoing programs.

Overall CGF Summary Statistics - Completed						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	28	2.83	1.37	3.69	2.91	1.30
1980	41	1.50	0.98	2.15	1.31	1.43
1990	29	1.41	1.01	1.00	1.27	0.71
2000	18	1.14	1.11	0.44	0.25	0.39

Table 8. Overall CGF summary statistics – Completed programs.

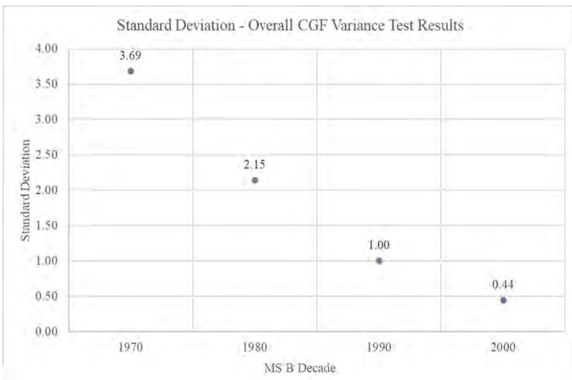

Figure 3. Standard deviations - Overall CGF of completed programs.

Ongoing

This section analyzes just the ongoing MDAPs. The 2000s and 2010s contained the bulk of our ongoing programs, but there are a couple of programs from the 1980s and 1990s that are still active and ongoing (e.g., reporting development/ production SARs). Because the K-W test needs at least five observations per group for statistical validity (Kruskal & Wallis, 1952), we removed from consideration any decade that did not meet the sample size criteria for either the schedule, total program CGF, or PAUC CGF analysis.

Table 10 presents the descriptive statistics for schedule growth by decade. As we have seen previously, both the means and medians are positive indicating consistent schedule slippage throughout the years for a typical MDAP program. The K-W and B-F tests returned p-values of 0.1067 and 0.9398, respectively, indicating no statistical difference among the decades with respect to the amount of schedule growth. This conclusion has been consistent throughout our analysis.

Table 11 presents the total program CGF for decades that had five or more MDAPs reporting development/production SARs. The K-W and B-F returned p-values of 0.0069 and 0.0020, respectively. The p-value of 0.0069 suggests that the decades are different with respect to total program CGF. The S-D test returned a p-value of 0.0096 when comparing the 1990s and 2010s, indicating that the 1990s total program CGF were

PAUC CGF Summary Statistics - Completed (Excluding C130 AMP)						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	25	1.55	1.36	0.91	1.10	0.59
1980	31	2.03	1.05	2.83	0.9	1.4
1990	29	2.29	1.26	2.88	1.37	1.26
2000	14	1.19	1.07	0.66	0.19	0.56

Table 9. PAUC CGF summary statistics – Completed programs (excluding C130 AMP).

Schedule Growth Percentage Summary Statistics - Ongoing MDAPs						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
2000	22	30%	39%	0.39	0.31	1.29
2010	23	14%	32%	0.32	0.30	2.29

Table 10. Schedule growth percentage summary statistics – Ongoing programs. Means and medians converted to percentages.

statistically higher than those of the 2010s. The 2000s were statistically equivalent to both decades. The 0.0020 p-value for the B-F test suggests that the standard deviations associated with total program CGF is not equal across the decades. As seen in Figure 4, there appears to be a decreasing trend in total program CGF variability by decade; a trend we witnessed in Figures 1 and 3.

Table 12 presents summary statistics of the PAUC CGF by decade with one outlier removed from the 1990s, the National Security Space Launch (NSSL) MDAP. This program possessed approximately a 6 PAUC CGF, while the next highest was around 1.6. The K-W and B-F tests returned p-values of 0.2564 and 0.0001, respectively. The p-value of 0.2564 suggests PAUC CGF through the three decades investigated are statistically equivalent. The p-value for the B-F test suggests that the variability associated with PAUC CGF is not equal. As seen in Figure 5, there appears to be a decreasing trend in PAUC CGF variability for the last three decades; a trend also shared by total program CGF. The next section discusses the significance of the statistical findings from our analysis.

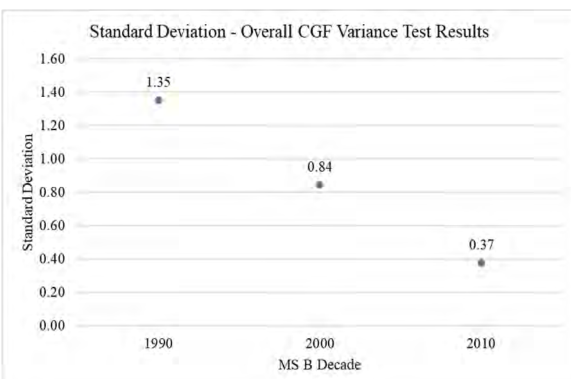

Figure 4. Standard deviations - Overall CGF of ongoing programs.

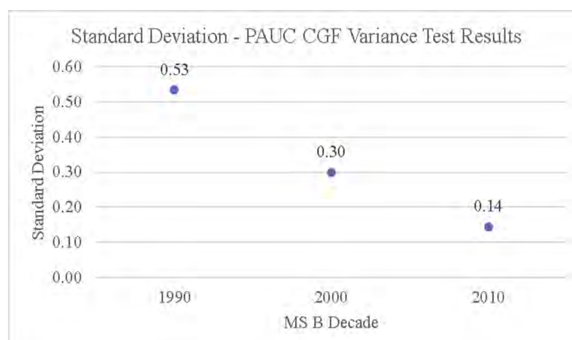

Figure 5. Standard deviations - PAUC CGF of ongoing programs.

Overall CGF Summary Statistics - Ongoing MDAPs						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1990	8	2.56	2.79	1.35	2.24	0.53
2000	24	1.47	1.21	0.84	0.64	0.57
2010	27	1.16	1.04	0.37	0.24	0.32

Table 11. Overall CGF summary statistics – Ongoing programs.

PAUC Summary Statistics - Ongoing MDAPs (Excluding NSSL)						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1990	7	1.00	1.21	0.53	0.91	0.54
2000	24	1.16	1.15	0.30	0.44	0.26
2010	27	1.01	1.02	0.14	0.19	0.14

Table 12. PAUC CGF summary statistics – Ongoing programs (Excluding NSSL).

Discussion and Conclusion

This article investigated whether cost and schedule estimations are improving over the decades. Despite numerous reforms and initiatives enacted to improve cost and schedule performance, our analysis found very few instances where schedule growth, total program CGF, or PAUC CGF statistically differed across the decades. Rather, our finding corroborated previous studies such as Arena et al. (2006) and Younossi et al. (2007) where schedule and cost growth are consistently positive across the decades.

Although the initial purpose of this study was to examine average cost and schedule trends, the most novel and exciting results were found elsewhere. This novel finding was found through an examination of the standard deviations of the CGFs across the decades. As shown throughout the analysis, even when the CGFs themselves were not statistically different across decades, there were differences detected in the variances of the CGFs themselves. This observation was seen for both overall CGF and PAUC CGFs for ongoing and completed programs. Perhaps most exciting is that these variances were generally decreasing. The overall CGF variance decreased through the five decades reviewed, while the PAUC CGF variance has decreased in every decade since the 1980s. Similarly, for completed programs the overall CGF variance has decreased since the 1970 while the PAUC CGF variance of on-going programs has decreased since the 1990s. See Figure 6.

To reiterate, although there were no identifiable statistical trends pointing to the DoD improving its schedule or cost estimation accuracy, the variances of the cost estimates have been noticeably decreasing from the 1980s onward. This decrease appears to be a new finding not seen in the literature previously. MDAPs are very expensive and time-consuming programs. Frequently, these programs are pushing technology capabilities. That alone suggests that cost growth and schedule slippage might just be endemic to MDAPs. However, the decreasing variability of cost estimates suggest to us that

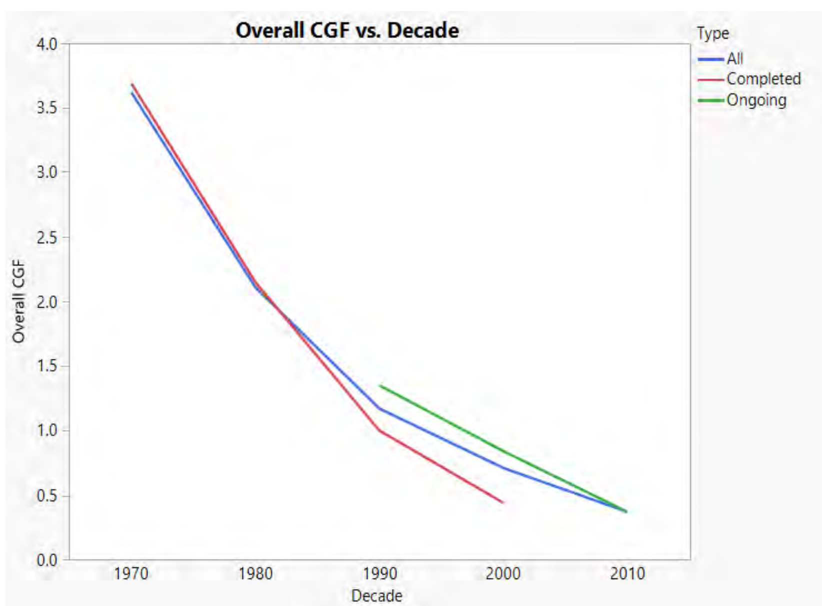

Figure 6. Standard deviations of overall total program CGFs.

cost estimators and/or the process behind them might be improving over time.

In statistics, those combinations speak to a bias outcome with minimal variance. In our opinion based on our experience in analyzing MDAPs over the years, that suggests perhaps the continued systematic bias of keeping initial cost estimates on the smaller side to make budgets more palatable. But eventually, the inherent risks of MDAPs are realized and true costs start accumulating. That is when cost growth appears. However, the variability of this cost growth difference has been reducing over the decades. That is the good news story. This is not to say that cost estimating cannot improve. However, we believe this cost growth is more of an artifact of keeping cost appearances low and not a reality of poor cost estimating. Thus, we humbly suggest program managers and executives consider this information when selecting the confidence level for budgetary inputs from a MS B estimate.

References

Acquisition Category. (2021). In AcqNotes [Online encyclopedia]. Retrieved from https://acqnotes.com/acqnote/acquisitions/milestone-b

Arena, M. V., Leonard, R. S., Murray, S. E., & Younossi, O. (2006). *Historical cost growth of completed weapon system programs* (Report No. TR-343). Santa Monica, CA: RAND. https://apps.dtic.mil/sti/pdfs/ADA455020.pdf

Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, *69*(346), 364–367. https://doi.org/10.1080/01621459.1974.10482955

Defense Industry Daily. (2014). Retrieved from https://www.defenseindustrydaily.com/c-130-amp-program-receives-contract-heads-into-uncertainty-06046/#:~:text=The%20C%2D130%20AMP%20program%20was%20originally%20supposed%20to%20cost,come%20under%20a%20separate%20contract.

Dwyer, M., Tidwell, B., Blivas, A., & Hunter, A. (2020, April 8). *Cycle times and cycles of acquisition reform*. Acquisition Research Program, Naval Postgraduate School. https://dair.nps.edu/bitstream/123456789/4183/1/SYM-AM-20-046.pdf

Fox, J. R., Allen, D., Moody, W., & Shiman, P. (2011). *Defense acquisition reform 1960-2009: An elusive goal.* U.S. Government Printing Office. https://history.defense.gov/Portals/70/Documents/acquisition_pub/CMH_Pub_51-3-1.pdf

Government Accountability Office. (2021, June). *Weapon systems annual assessment: Updated program oversight approach needed* (GAO-21-222). United States Government Accountability Office. https://www.gao.gov/assets/gao-21-222.pdf

Kozlak, S., White, E., Ritschel, J., Lucas, B., & Seibel, M. (2017). Analyzing cost growth at program stages for DoD aircraft. *Defense Acquisition Research Journal*, *24*(3), 386–407. https://doi.org/10.22594/dau.16-763.24.03

Kruskal, W.H. & Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*, 583-621.

Laerd Statistics. (2018). *Kruskal-Wallis H Test in SPSS Statistics | Procedure, output and interpretation of the output using a relevant example.* Lund Research Ltd. https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php

Light, T., Leonard, R. S., Pollak, J., Smith, M. L., & Wallace, A. (2017). *Quantifying cost and schedule uncertainty for Major Defense Acquisition Programs (MDAPs)* (Report No. RR1723). Santa Monica, CA; RAND. https://www.rand.org/pubs/research_reports/RR1723.html

Major Defense Acquisition Program. (2020). In AcqNotes [Online encyclopedia]. Retrieved from https://acqnotes.com/acqnote/acquisitions/major-defense-acquisition-program

McNicol, D (2018). *Acquisition policy, cost growth, and cancellations of Major Defense Acquisition Programs* (IDA Publication No. R-8396). Institute for Defense Analysis. https://apps.dtic.mil/sti/pdfs/AD1122550.pdf

Monaco, J. V. & White, E. D. (2005). Investigating schedule slippage. *Defense Acquisition Review Journal*, *12*(2), 176–193. https://doi.org/10.21236/ada441770

Nunn-McCurdy Breach. (2021). In DAU Glossary [Online encyclopedia]. Retrieved from https://www.dau.edu/glossary/Pages/GlossaryContent.aspx?itemid=28040

Riposo, J., McKernan, M., & Kaihoi, .Chelsea. (2014). *Cycle times and schedule growth in defense acquisition: A literature review* (RR-455-OSD). Santa Monica, CA: RAND. http://www.rand.org/pubs/research_reports/RR455.html

Stephanie, G. (2015, March 11). *Brown-Forsythe test: Definition*. Statistics How To. Retrieved December 15, 2021, from https://www.statisticshowto.com/brown-forsythe-test/

Under Secretary of Defense, Acquisition, Technology, and Logistics (USD(AT&L). (2016). *Performance of the Defense Acquisition System: 2016 Annual Report*. Washington, DC: Department of Defense. Available at https://apps.dtic.mil/sti/pdfs/AD1019605.pdf

Younossi, O., Arena, M. V., Leonard, R. S., Roll, C. R., Jain, A., & Sollinger, J. M. (2007). *Is weapon system cost growth increasing? A quantitative assessment of completed and ongoing programs* (Report No. MG-588-AF). Santa Monica, CA: RAND. https://doi.org/10.7249/mg588

---

*Captain **Sammantha Jones**, USAF, is a recent graduate of the Air Force Institute of Technology (AFIT) where she earned a Master's of Science in Cost Analysis. She is currently assigned as a cost analyst for the Air Force Life Cycle Management Center Enterprise IT and Cyber Infrastructure Division at Hanscom AFB. (E-mail address: sammantha.jones@us.af.mil)*

*Dr. **Edward D. White** is a Professor of Statistics in the Department of Mathematics and Statistics at AFIT. He received his BS in Mathematics from the University of Tampa, MAS from The Ohio State University, and PhD in Statistics from Texas A&M University. His primary research interests include statistical modeling, simulation, and data analytics. (E-mail address: Edward.White@afit.edu)*

*Dr. **Jonathan D. Ritschel** is an Associate Professor of Cost Analysis in the Department of Systems Engineering and Management at AFIT. He received his BBA in Accountancy from the University of Notre Dame, MS in Cost Analysis from AFIT, and PhD in Economics from George Mason University. His research interests include public choice, the effects of acquisition reforms on cost growth in DoD weapon systems, and economic institutional analysis. (E-mail address: Jonathan.Ritschel@afit.edu)*

*Mr. **Shawn M. Valentine** is an Operations Research Analyst at AFLCMC/FZCE at Wright Patterson AFB, Ohio. Mr. Valentine graduated from Ohio University with a Bachelor of Science (BS) in Actuarial Science and a Master of Science (MS) in Financial Economics. (Email address: shawn.valentine@us.af.mil)*

The International Cost Estimating and Analysis Association is a 501(c)(6) international non-profit organization dedicated to advancing, encouraging, promoting and enhancing the profession of cost estimating and analysis, through the use of parametrics and other data-driven techniques.

www.iceaaonline.com

Submissions:

Prior to writing or sending your manuscripts to us, please reference the JCAP submission guidelines found at

www.iceaaonline.com/publications/jcap-submission

Kindly send your submissions and/or any correspondence to JCAP.Editor@gmail.com

International Cost Estimating & Analysis Association

4115 Annandale Road, Suite 306 | Annandale, VA 22003

703-642-3090 | iceaa@iceaaonline.org