



11100 Johns Hopkins Road • Laurel, Maryland 20723-6099

Advanced Earned Value Management: Using Time Series Forecasts and Regression Models

2024 ICEAA Professional Development & Training Workshop

Anna B. Peters, Johns Hopkins University Applied Physics Lab
Mark W. Hodgins, Ph.D., Johns Hopkins University Applied Physics Lab
Adam J. Levin, Johns Hopkins University Applied Physics Lab

Distribution Statement A. Approved for public release: distribution is unlimited. (DoD 5230.24, 2023)

Abstract

The recent digitization of contractor EVM data affords cost analysts a newfound ability to execute robust statistical and data science techniques that better predict total project cost and schedule realism. Time series analysis, a well-established method in private sector finance, is one such method. Auto regressive integrated moving average (ARIMA) models may capture the persistence and patterns in EVM data, as measured by CPI, SPI, and schedule execution metrics (SEMs). As a second option, macroeconomic regression models can measure the relationship between contract performance and external considerations, like unemployment and inflation, over time. Both techniques, moreover, may forecast future changes in EVM variables interest, like IEAC. This presentation will discuss how these types of time series models and forecasts are employed on real acquisition programs and their associated IPMDAR data using Python based tools to raise program analysts' alertness to emergent acquisition risks and opportunities.

Abstract..... ii

1. Introduction..... 1

2. Analytic Enablers 1

3. Time Series Analysis..... 4

 3.1 Analytical Background..... 4

 3.2 ARIMA using NRO SEMs..... 6

 3.3 ARIMA using CPI and SPI..... 9

4. Macroeconomic Regression..... 15

5. Path Forward..... 18

6. Conclusions 19

List of Figures

FIGURE 1. CURRENT BEI MODEL, FOR NON-CUMULATIVE VIEW	7
FIGURE 2. CURRENT BEI MODEL, 3 REPORTING PERIOD CUMULATIVE VIEW	8
FIGURE 3. CURRENT BEI MODEL, 6 REPORTING PERIOD CUMULATIVE VIEW	8
FIGURE 4. COMPARING THE 6 REPORTING PERIOD CUMULATIVE VIEWS OF BRI, BPI, AND SPI.....	9
FIGURE 5. CUMULATIVE CPI MODEL, FOR WBS ELEMENT 1.5.1.....	12
FIGURE 6. CUMULATIVE SPI MODEL, FOR WBS ELEMENT 1.5.1.....	13
FIGURE 7. CUMULATIVE CPI MODEL, FOR WBS ELEMENT 1.5.1.1.....	13
FIGURE 8. CUMULATIVE SPI MODEL, FOR WBS ELEMENT 1.5.1.1.....	14
FIGURE 9. MEASURES OF ECONOMIC PERFORMANCE.....	16
FIGURE 10. UNEMPLOYMENT AND CUMULATIVE CPI	18

LIST OF TABLES

TABLE 1. OUTLINES THE MAJOR ELEMENTS OF THE DATA ENVIRONMENT	3
TABLE 2. MAJOR PYTHON LIBRARIES USED IN THIS ANALYSIS.....	4
TABLE 3. COMPARING DESCRIPTIVE IEAC AND BAC FOR THIRD LEVEL WBS ELEMENTS.....	14
TABLE 4. COMPARING DESCRIPTIVE IEAC AND BAC FOR FOURTH LEVEL WBS ELEMENTS.....	15

1. Introduction

Earned value management (EVM) is a best practice that Department of Defense (DoD) acquisition programs implement to monitor cost and schedule requirements and performance. However, the implementation of these analytic principles is not a panacea for effective cost and schedule performance monitoring as large-scale DoD acquisition programs routinely experience cost overruns and schedule delays (Meier, 2010). Government Accountability Office (GAO) has completed several studies comparing the frequency and severity of cost and schedule overruns in large acquisition programs occurring in 2000 vs 2008 and has found a higher rate of cost and schedule overruns in 2008 than 2000 (Meier, 2010). Specifically, Meier (2010) determined that, based on several GAO study results in this vein, there was “in absolute terms, a 702% cost growth increase” and a “31% increase in schedule delays” over the 7-year period considered.

Fortunately, several new data sources, techniques, and metrics have created an opportunity to apply EVM analysis using advanced data science techniques to assist in the prevention of cost and schedule overruns. As an example, the relatively new requirement for large DoD acquisition programs to electronically publish an integrated program management data and reports (IPMDARs) each reporting period provides the analyst with much higher fidelity data relative to non-digital Integrated Program Management Data and Analysis Reports (IPMRs) (IPMDAR Contract Performance, Schedule Performance, Implementation, 2020). Moreover, the evolution of data science platforms provides the cost community with low-to-no cost tools and techniques to examine cost information. Finally, some newer advanced schedule metrics created by the National Reconnaissance Office (NRO), named the schedule estimating metrics (SEMs), are becoming another robust option to track a program’s schedule performance.

Given these opportunities, this study presents how relatively advanced statistical techniques are applied to acquisition data in order to improve the cost and schedule performance forecasting. Prior to discussing the specific methodological implementations, Section 2 examines the recent data source, technique, and metric opportunities in greater detail. Section 3 then describes the implementation of time series analysis forecasting on acquisition data sets. The study team executes the ARIMA technique to forecast future values of standard EVM, like cost performance index (CPI), and schedule performance index (SPI), and future measures of the NRO SEMs. External factors may also affect DoD acquisition performance. Section 4 executes an analytic process using linear regression with lags to explore how macroeconomic indicators, such as unemployment and inflation rates, may influence acquisition health. Section 5 and 6 explore future research and conclude the study, respectively.

2. Analytic Enablers

The application of new statistical and data science approaches to forecast acquisition performance is enabled in this effort by emergent metrics, data sources, and techniques. Specifically, the NRO SEMs provide additional descriptive metrics to conduct time series analysis, IPMDARs represent a new data set with more detailed contract performance information, and data science tools and

techniques enable robust analysis. Each of these opportunities is discussed in detail throughout the remainder of this section

New data metrics are one key enabler. The NRO created eight advanced schedule analysis techniques, referred to as SEMs, to provide objective and improved measures of schedule performance (NRO, 2022). Specifically, NRO leadership wanted the metrics to serve as advanced predictions of inadequate schedule accomplishment before threshold levels are violated and costly re-programming is required (Schultheis, 2021). The NRO study team used Naval Air Systems Command (NAVAIR) data to derive their methods and test metric reliability on a portfolio of completed projects (Schultheis, 2021).

This analysis focused on three of the eight SEMs to support acquisition performance forecasting:

1. Current Baseline Realism Index (BRI) – Percentage of planned events that actually finished in the planning period; indicator of how well the contractor is following the plan in the period. NRO recommends BRI is applied using a six-reporting period moving average. (NRO, 2021)
 - $Current\ BRI = \frac{\sum\ of\ tasks\ baselined\ to\ finish\ in\ RP\ \&\ completed\ in\ RP}{\sum\ tasks\ baselined\ to\ finish\ in\ RP}$
 - $BRI \geq 0.80$ considered favorable and “on plan”. There are no major milestone delays or baseline resets expected in next 6-12 months.
 - $BRI \leq 0.20$ considered unfavorable and “way off plan”.
2. Current Baseline Progress Index (BPI) – Percentage of planned events that actually finished in or before the planning period; indicator of how many of the planned events in the period have actually been accomplished. (NRO, 2021)
 - $Current\ BPI = \frac{\sum\ of\ tasks\ baselined\ to\ finish\ in\ RP\ \&\ completed\ in\ or\ before\ RP}{\sum\ tasks\ baselined\ to\ finish\ in\ RP}$
 - $BPI \leq 0.35$ considered unfavorable and “way off plan”.
3. Current Baseline Execution Index (BEI) – Percentage of total events that actually finished in the current period; pace of work. (NRO, 2021)
 - $Current\ BEI = \frac{\sum\ of\ tasks\ finished\ in\ RP\ even\ if\ not\ originally\ planned\ to\ finish\ in\ RP}{\sum\ tasks\ baselined\ to\ finish\ in\ RP}$

The favorable and unfavorable thresholds for BRI and BPI were estimated from a statistical analysis of historical data from completed government acquisition programs (NRO, 2021). Programs actual acquisition performance was used to classify the favorable and unfavorable levels based on realized issues and positive impacts. Due to the data driven methods to derive the metrics and the associated classifiers, data science and statistical techniques that employ the descriptive SEMs in a predictive manner may provide new and improved analytic insights. In fact, JHU APL and NRO are collaborating to further advance predictive analytics to improve acquisition program decision making and forecasting.

IPMDARs are a second enabler. Data Item Description (DID) DI-MGMT-81861C specifies the attributes of IPMDARs and contractually requires their use on significant DoD acquisition

efforts (e.g., typically in excess of \$20 million) (DID for IPMDAR, 2021). IPMDARs include three primary data components, which are the contract performance dataset (CPD), the schedule performance dataset (SPD), and the performance narrative report (IPMDAR Contract Performance, Schedule Performance, Implementation, 2020). The CPD includes most of the earned value information and the program baseline data. The SPD contains the contractor's integrated master schedule (IMS), and the narrative reports provide supporting documentation to accompany the CPD and SPD.

Critically, IPMDARs possess instructions to ensure formatting rigor and maturity. File format specifications (FFSs) and data exchange instructions (DEIs) provide precise digitization instructions for transparency, replicability, and auditability for the CPD and SPD. In addition, the submittals are required on a monthly basis. The digitized data, reliable formatting, and frequency of reporting afford the ability to execute time series analysis on major development contracts. In turn, the newly available IPMDAR data for current and emergent acquisition programs provides an opportunity to improve upon the standard EVM analysis limitations.

The data environment is a third and final enabler for advanced statistical analysis. Low-to-no cost data science environments are now common and a burgeoning supply of data scientists with the requisite human capital execute complex analytics within these environments. Python is one such environment, which the study team selected for executing analytic techniques across the data pipeline. Python is highly used and a top ranked data science programming language, with significant community support. Python programming is relatively accessible because this open-source language possesses highly readable and writable syntax/code. Finally, Python supports a variety of emergent, predictive analytics on disparate sized relational data sets, which aligns with the study team's requirements.

Overall, the environment is a multi-tiered (e.g., n-tier), where the first tier is the database and the schemas that store the structured data. A benefit of this architecture is the ability to process the information in any desired manner (e.g. data normalization or transformations). The second tier is the business logic, which is the analytics tier in this case. The business logic has access to the database through a customized python library via SQL queries and is supported by the psycopg2 Python library. The business logic tier is where the time series and macroeconomic regression analysis are conducted. The third tier is the presentation layer, which presents the analytics' visualizations using Jupyter notebooks and a plotly-dash dashboard (Plotly API, 2023). Table 1 outlines the major elements of the data environment. Table 2 lists the major Python libraries used across all analysis discussed in this report.

Table 1. Outlines the major elements of the data environment

Item Type	Description
Database (DB)	SQL (relational) database
DB management system	PostgreSQL
SQL toolkit and object-relational mapper	SQLAlchemy
PostgreSQL DB adapter	Psycopg2 (2.9.5)

Software (Programming Language)	Anaconda (Python)
---------------------------------	-------------------

Table 2. Major Python libraries used in this analysis

Python Library	Purpose
ARIMA, from statsmodels.tsa.arima.model	ARIMA time series modeling
auto_arima, from pmdarima.arima	Time series analysis, specifically auto_arima
dateutil, from relatedelta	Managing date-time formatted data
express, from plotly	Data visualization
graph_objects, from plotly	Data visualization
matplotlib.pyplot	Creating object-oriented plots
numpy	Scientific computing
pandas	Building and manipulating data structures
product, from itertools	Creating iterators for efficient looping
seaborn	Creating statistical data visualizations

3. Time Series Analysis

3.1 Analytical Background

Time series modeling serves as a focal point of this effort due to the enabling data, metrics, and tools discussed in the previous section. In general, time series analysis is a statistical method used to determine potential trends in a data set over time. A statistical background is briefly discussed here before turning to how this analysis can be applied to EVM.

In time series modeling, time is an independent variable that is used to evaluate the potential for statistically significant relationships with a dependent variable (Cryer, et. al., 2008). The time period or frequency can vary with each time series model, allowing for different relationships to be investigated. Autoregressive integrated moving average (ARIMA) is one of the most common forms of time series analysis. ARIMA has been used extensively in the financial industry to forecast stock prices over time. ARIMA is also used frequently in epidemiology to predict the spread of disease over time.

The ARIMA model is a combination of two other time series model formats: the auto regressive (AR) model and the moving average (MA) model (Cryer, et. al., 2008). The AR model is constructed so that Y depends only on its own lags (Cryer, et. al., 2008). The MA model is constructed so that Y depends on the lagged forecast errors (Cryer, et. al., 2008). The integration

of the two model types is done once the stationarity assumption has been validated. The stationarity assumption can be validated in a variety of ways, in this analysis the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) unit-root test was applied (Kwiatkowski, et. al., 1992). ARIMA modeling can be applied using seasonal or non-seasonal trends. This analysis focuses on non-seasonal ARIMA modeling, due to acquisition programs and the IPMDAR data not having seasonal tendencies. The theoretical ARIMA equation is below, Equation 1, where Y_{t-p} and $\phi_q \epsilon_{t-q}$ are the Autoregressive (AR) and Moving Average (MA) components of the model, respectively (Cryer, et. al., 2008):

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_q \epsilon_{t-q}$$

Equation 1. Theoretical ARIMA

The more commonly used and abbreviated notation for a non-seasonal ARIMA model is: ARIMA(p, d, q) (Cryer, et. al., 2008). The ARIMA(p, d, q) model is comprised of three terms: p, d, and q. P is the order of the AR term (the number of lags of Y to be used as predictors), d is the minimum number of differencing needed to make the time series stationary (if the series is already stationary than d=0), and q is the order of the MA term (number of lagged forecast errors that should go into the ARIMA model) (Cryer, et. al., 2008).

Critically, ARIMA models can also be used to forecast future trends based on historical data to predict what the future values of the dependent variable could be (Cryer, et. al., 2008). In this analysis, one of the IPMDAR metrics (e.g. CPI, SPI, and SEMs) are used as the dependent variable to build an ARIMA model. That model will then be built upon to forecast future CPI, SPI, IEAC, BRI, BPI, and BEI values (Montgomery, 2015).

When building the framework to forecast future metric values this analysis followed the best practice of separating the dataset into test and validation subsets. The test data makes up 80% of the original dataset and is used to create models to explain the data available. The remaining 20% of the data is put into a validation subset, which is used to verify the models chosen using the test data set. Time series analysis requires a continuous time series to use as the independent variable, so the data was separated based on the first 80% in time as the test dataset and the last 20% in time as the validation dataset. After validating that the models adequately explain the data, they are applied to the entire dataset.

Confidence intervals were added to each metric's forecast window to account for the level of uncertainty inherent in any modeling forecast. The confidence intervals display what range the actual metric results could have based on the forecasted point estimate. The higher the confidence interval percentage, the wider the range of data considered as part of the uncertainty bounds for each forecast point estimate. This analysis displays a 90%, 80%, and 50% confidence interval band for all forecast results. The varying levels are employed to reflect varying risk preferences among decision makers.

The specific ARIMA(p, d, q) values for each model were chosen using a python function called 'auto_arima' to optimize the process of determining the best model. The 'auto_arima' function automatically checks the stationarity assumption using the KPSS test and then iterates through possible combinations for the best ARIMA model. The final model is chosen based on which

iteration has the lowest value of the Akaike information criterion (AIC) (Keaton, 2011). After finalizing the best ARIMA models, those models were used to forecast the dependent variable values up to six reporting periods into the future. Forecasts are calculated six reporting periods into the future to assist in the next rolling wave of EVM planning that acquisition programs typically complete every six months. The forecasted values are based on the historical dependent variable values that are contained within the model.

3.2 ARIMA using NRO SEMs

The study team used the ARIMA method to build time series models for the BRI, BPI, and BEI SEMs. Time was used as the independent variable in each model. The dependent variable for each model was either the BRI, BPI, and BEI values. These relationships were also used to forecast or extrapolate significant trends to future time periods. In this analysis, ARIMA can provide insight to address how the SEMs are predicted to change in the near future and if the forecasts are on track with previously defined SEM values.

In order to model the SEM BRI, BPI, and BEI metrics effectively, independent non-seasonal ARIMA models were used. The dependent variable metrics used in the ARIMA models are calculated in several different ways. The study team calculated several variants for each dependent variable based on time durations. While the BRI calculation is limited to a 6-reporting period moving average, BPI and BEI were estimated for each reporting period using 1 reporting period of data (current time and non-cumulative), 3 reporting periods of cumulative data, and 6 reporting periods of cumulative data. Creating three models for BPI, and BEI allows for a more comprehensive look at the BPI, and BEI values because each model reacts differently to the historical data. Typically, an ARIMA model using a SEM calculated with 1 reporting period of non-cumulative data is the most reactive to changes in the data, an ARIMA model using a SEM calculated with 3 reporting periods of cumulative data is a bit smoother and less reactive than the one reporting period model, and finally an ARIMA model using a SEM calculated with 6 reporting periods of cumulative data tends to be the least reactive and most conservative of the three modeling views.

The three model views of BEI are presented below to demonstrate the difference in the model and forecasts chosen with each view. Finally, the 6-reporting period moving average view of BRI and 6-reporting period cumulative views of BPI, and SPI were also compared to demonstrate the difference between the new SEMs and traditional EVM in identifying program schedule risk. The reader must note that all analytic visualizations presented are performed on notional data and intended to serve as illustrations of how to execute the analysis and interpret the results. However, the general results are reflective of trends observed when applying the analysis to several acquisition programs.

The three BEI models, which are the 1 reporting period (Figure 1), 3 reporting period cumulative (Figure 2), and 6 reporting period cumulative (

Figure 3), reflect the differences in the responsiveness of the metric based on how much historical data is used to build the dependent variable. As predicted, the non-cumulative view is the most reactive and the 6-reporting period cumulative view is the most conservative with the smoothest trendline fit. Overall, all three BEI models show a downward trend indicating that there are likely schedule risks happening within the program. The BEI forecasts for each model mostly follow this general downward trend as well.

Figure 4 compares the 6-reporting period moving average view of BRI and the 6-reporting period cumulative view for BPI against a traditional EVM metric, SPI. BPI predicts that the program will be “way off plan” starting in reporting period 22. BRI predicts that the program will be “way off plan” in reporting period 35. The results of the BRI model and forecasts indicate that each month the percentage of completed tasks that were baselined to be executed has reduced; the schedule issues are forecasted to persist despite the recent improvement. The results of the BPI model and forecasts indicated that, similarly to BRI, the percentage of completed tasks that were baselined to be executed leading up to the current reporting period reduces; the schedule issues are forecasted to persist. On the contrary, the SPI values are not indicative of significant program schedule risk. The SPI values briefly drop below the 0.8 mark but quickly return to a reasonable SPI range, indicating there is not a lingering program schedule risk. As a result, the SEMs are more responsive to schedule issues. In this example, BRI forecast the program will be “way off plan” starting in reporting period 35 and BPI detects the program to be “way off plan” starting in reporting period 22 whereas SPI is expected to trend right above the 0.8 threshold and not raise alarm of a major schedule issue. Implementation of the SEMs may allow program personnel more time to address these issues before it potentially became a critical issue.

Figure 1. Current BEI model, for non-cumulative view

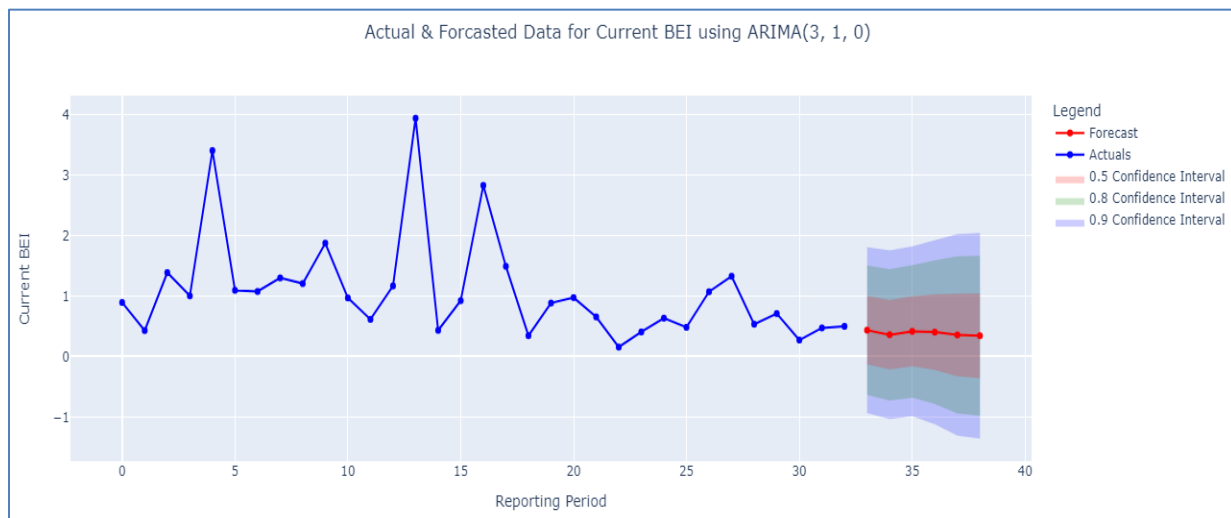


Figure 2. Current BEI model, 3 reporting period cumulative view

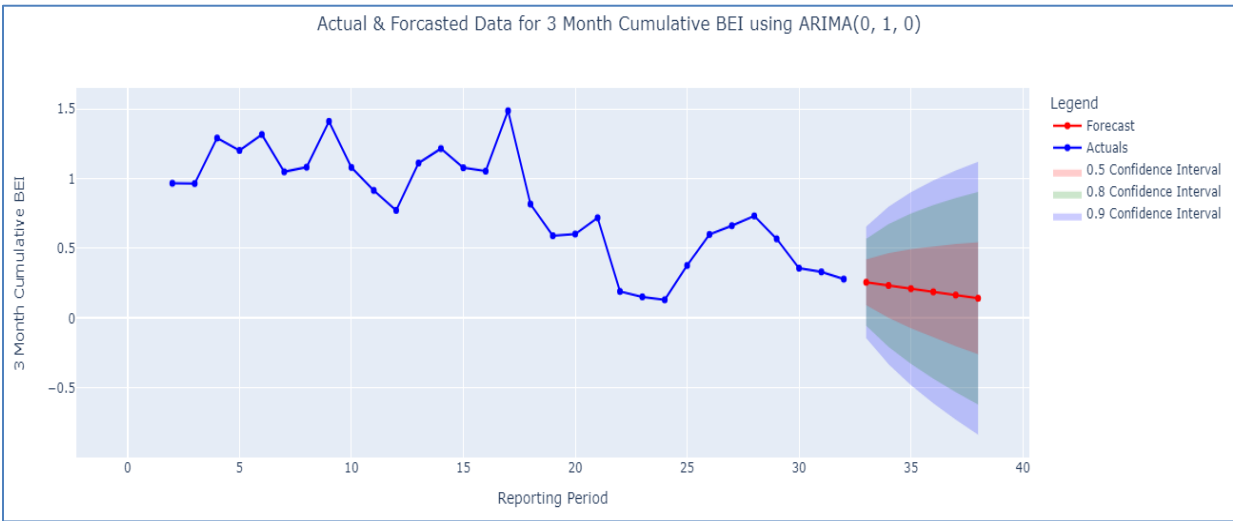


Figure 3. Current BEI model, 6 reporting period cumulative view

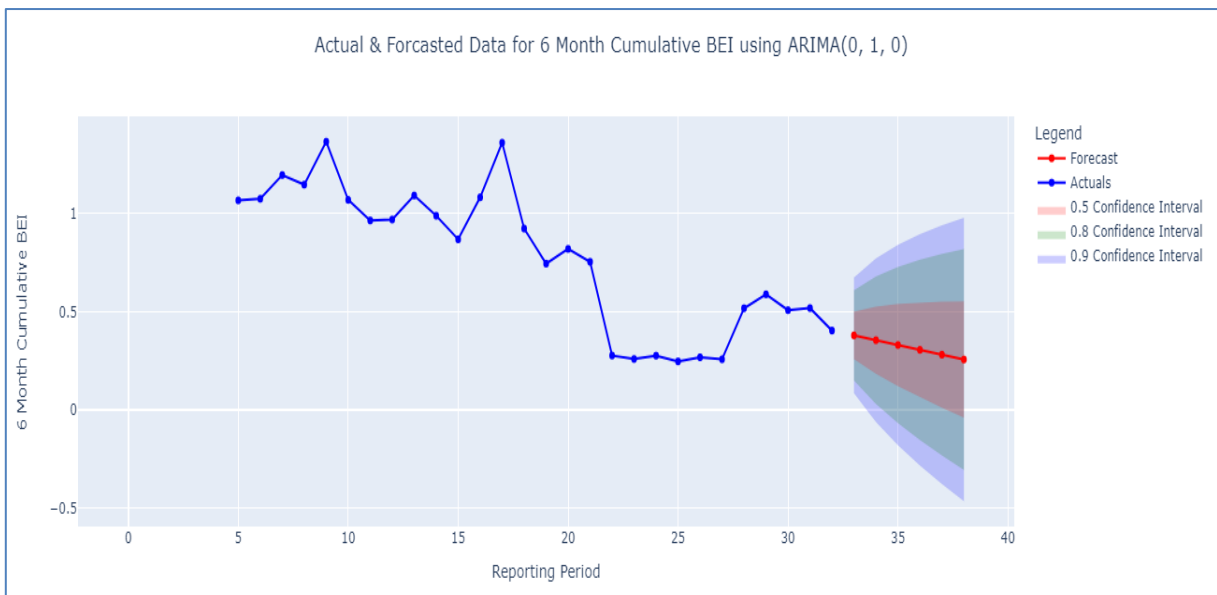
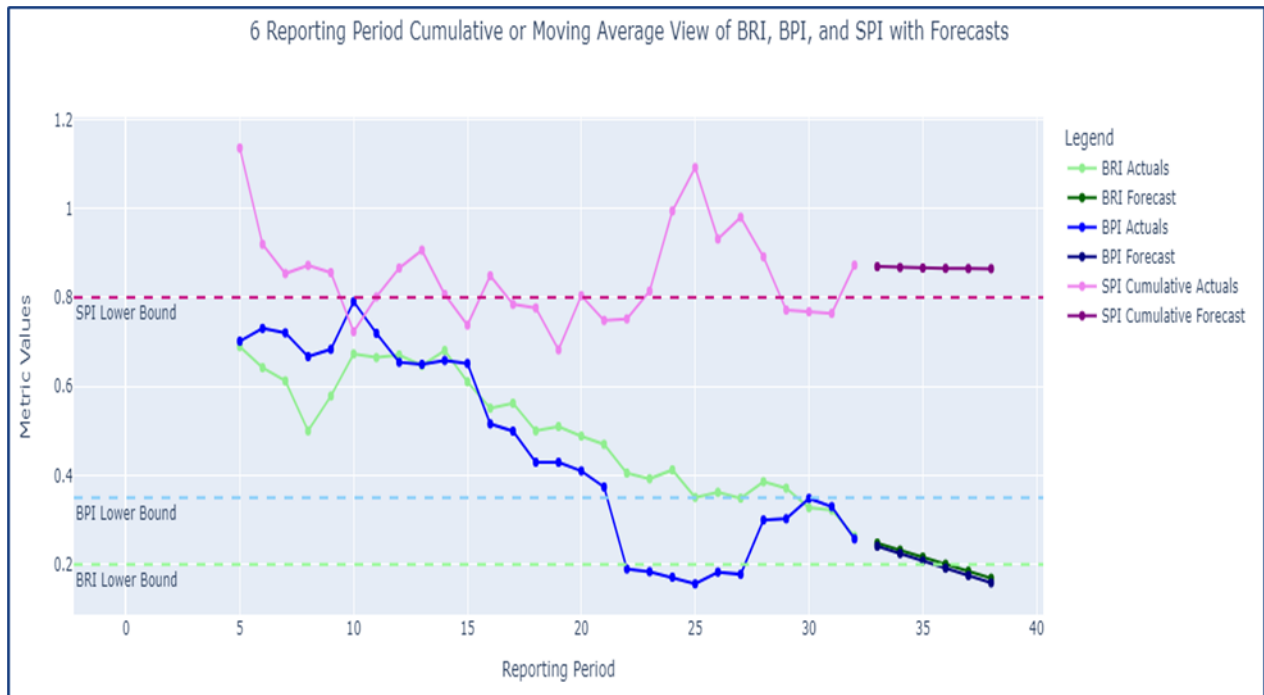


Figure 4. Comparing the 6 Reporting Period Cumulative Views of BRI, BPI, and SPI

3.3 ARIMA using CPI and SPI

The study team also executed ARIMA modeling to forecast traditional EVM metrics, like cumulative SPI and cumulative CPI, at varying levels of the work breakdown structure (WBS). In this case, the ARIMA models use time as the independent variable and the CPI or SPI calculations, using the IPMDAR dataset, as the dependent variable. These relationships were also used to forecast or extrapolate significant trends to future time periods. In this analysis, ARIMA can provide insights on how CPI and SPI are estimated to change in the near future, if predicted CPI and SPI rates are on track with recent history, and how abnormal changes to earned value may affect future cost and schedule performance. Moreover, the capability to drill up and down through the WBS levels allows account managers to conduct root cause analysis with models that are updated and refined each reporting period.

An independent estimate at complete (IEAC) prediction is feasible as well. An IEAC prediction may improve upon the standard IEAC through the incorporation of cost and schedule forecasts from the ARIMA CPI and SPI analysis. For reference, the standard IEAC is used as a test for “reasonableness” for the contractor provided estimate at complete (EAC) (Keaton, 2011). IEAC can also be used as an external approximation of the budget at complete (BAC) (Keaton, 2011). BAC is a stable value that is a baseline budget planned for the entire program effort based on the amount of authorized funding. The BAC can change when a program re-baseline occurs. This

analysis will focus on the comparison of IEAC to BAC. Tracking the IEAC can help balance work priorities, re-plan remaining tasks, and adjust the technical approach to complete the project with the remaining resources (Keaton, 2011).

This notional analysis uses the MIL-STD-881F Appendix G Ground Vehicle Systems WBS structure (DoD, 2022). The study focuses on the analytic implementation and results for two example WBS elements. In order to emphasize the multidimensional capabilities of this model, one example is executed at the third level of the WBS structure, while the second example is presented at the fourth level of the WBS structure. The example third level WBS element is 1.5.1 Development Test and Evaluation (DoD, 2022). The example fourth level WBS element is 1.5.1.1 Cybersecurity Test and Evaluation (DoD, 2022).

The dependent variables chosen for these ARIMA models were the IPMDAR CPI and SPI metrics. These metrics are fundamental, descriptive EVM metrics. CPI is used to keep track of how the project cost is performing (Keaton, 2011). SPI is used to keep track of how the project schedule is performing (Keaton, 2011). Both metrics set a threshold of greater than one to be considered favorable performance and less than one to be considered unfavorable performance (Keaton, 2011). The study team forecasted CPI and SPI at varying WBS; however, the calculation of these indices at higher fidelity, subsystem and component levels, introduced data quality challenges. Accounting revisions and missing data, for example, led to normalization and statistical issues. In order to limit noise in the IPMDAR data caused by accounting revisions and missing data, the ARIMA models use rolling cumulative CPI and cumulative SPI measurements as the dependent variables (Equation 2).

$$Cumulative\ CPI_t = \sum_{i=1}^j \frac{BCWP_i}{ACWP_i} \quad Cumulative\ SPI_t = \sum_{i=1}^j \frac{BCWP_i}{BCWS_i}$$

where j = latest reporting period

$$Threshold: \begin{cases} \text{Favorable,} & > 1 \\ \text{Unfavorable,} & < 1 \end{cases}$$

Equation 2. Cumulative CPI and Cumulative SPI

Independent non-seasonal ARIMA models were used to model the IPMDAR cumulative CPI and cumulative SPI metrics effectively. The ARIMA models were used to forecast 6 future reporting periods of CPI or SPI values. The forecasted cumulative CPI and cumulative SPI are incorporated into a novel, predictive estimate of the IEAC to compare against BAC. There are several ways to solve for IEAC but Keaton (2011) provides a general equation (See Equation 3).

$$IEAC = \frac{ACWP + (BAC - BCWP)}{CPI_{cum} * SPI_{cum}}$$

where CPI_{cum} or SPI_{cum} is the cumulative value of CPI or SPI

Equation 3. IEAC

In this analysis the traditional descriptive IEAC metric was adapted to derive an IEAC metric that incorporates the results of the cumulative CPI and cumulative SPI ARIMA forecasts. The descriptive IEAC (Equation 4) is modified into a comprehensive valuation that allows for the insertion of the cumulative CPI and cumulative SPI 6 reporting period forecast (Keaton, 2011). The forecasted IEAC (Equation 5) replaces part of the comprehensive IEAC calculation with 6 reporting period forecasts from the cumulative CPI, and cumulative SPI ARIMA models to allow for a more robust predictor when compared to the program's BAC. This methodology created a novel IEAC equation (Equation 5) with three distinct parts. The left most part of the equation contains all costs performed to date, as represented by the ACWP. The second, middle, part contains the forecasted CPI and SPI metrics for 6 future reporting periods. Finally, the third part contains the costs planned by the program until the end of the contract, as represented by a calculation of the estimate to complete.

$$IEAC_{comp} = ACWP_{cum} + \frac{BAC_t - BCWP_{cum}}{CPI_{cum} * SPI_{cum}}$$

where $t = \text{time}$, $ACWP_{cum} = \left(\sum_{i=1}^{t-1} acwp_i \right)$, $BCWP_{cum} = \left(\sum_{i=1}^{t-1} bcwp_i \right)$,

$$CPI_{cum} = \left(\sum_{i=1}^{t-1} \frac{bcwp_i}{acwp_i} \right), SPI_{cum} = \left(\sum_{i=1}^{t-1} \frac{bcwp_i}{bcws_i} \right)$$

Equation 4. IEAC, Comprehensive

$$IEAC_t = ACWP_{cum} + \sum_{j=t+1}^{t+6} \left(\frac{\widehat{BCWS}_j * \widehat{SPI}_j}{\widehat{CPI}_j} \right) + \frac{BAC_t - BCWP_{cum} - \sum_{j=t+1}^{t+6} \widehat{BCWP}_j}{CPI_{cum} * SPI_{cum}}$$

where $t = \text{time}$, $ACWP_{cum} = \left(\sum_{i=1}^{t-1} acwp_i \right)$, $BCWP_{cum} = \left(\sum_{i=1}^{t-1} bcwp_i \right)$,

$$CPI_{cum} = \left(\sum_{i=1}^{t-1} \frac{bcwp_i}{acwp_i} \right), SPI_{cum} = \left(\sum_{i=1}^{t-1} \frac{bcwp_i}{bcws_i} \right)$$

\widehat{BCWS}_j = BCWS to complete within forecast window of ARIMA models,

\widehat{SPI}_j = forecasted value of SPI based on ARIMA models,

\widehat{CPI}_j = forecasted value of CPI based on ARIMA models

Equation 5. IEAC, at time t

The cumulative CPI and cumulative SPI models for WBS element 1.5.1 Development Test and Evaluation (Figure 5 and Figure 6) both show an overall negative trend in the program's cost and schedule performance, based on the notional data (DoD, 2022). The forecasts for both models predict that this trend is going to continue for the next six reporting periods. Despite the overall negative trend, the cumulative CPI model never dropped below 0.8 indicating that there is not a

high cost risk for the program based on the current notional data. However, cumulative SPI floats right around the 0.8 threshold starting in reporting period 7 indicating there is some schedule risk for the program based on current notional data. Nonetheless, the 80% and 90% confidence intervals for both models forecast that cost and schedule risk could increase, which may warrant increased monitoring and enforcement activities by the program office.

Similar to the previous two models, the cumulative CPI and cumulative SPI models for WBS element 1.5.1.1 Cybersecurity Test and Evaluation (Figure 7 and Figure 8) both show an overall negative trend in the program’s cost and schedule performance, again based on the notional data (DoD, 2022). The forecasts for both models predict that this trend is going to continue for the next six reporting periods. In both the cumulative CPI and cumulative SPI figures a drastic drop in cost and schedule performance, below 0.8, is shown starting in reporting period 22 the trend for both metrics then slightly increases in the remaining reporting periods but neither cumulative CPI or cumulative SPI are forecasted to return to a low risk level in the remaining reporting periods shown. These results indicate that preventive actions, like a root cause analysis, may be necessary by the program office.

Figure 5. Cumulative CPI model, for WBS Element 1.5.1

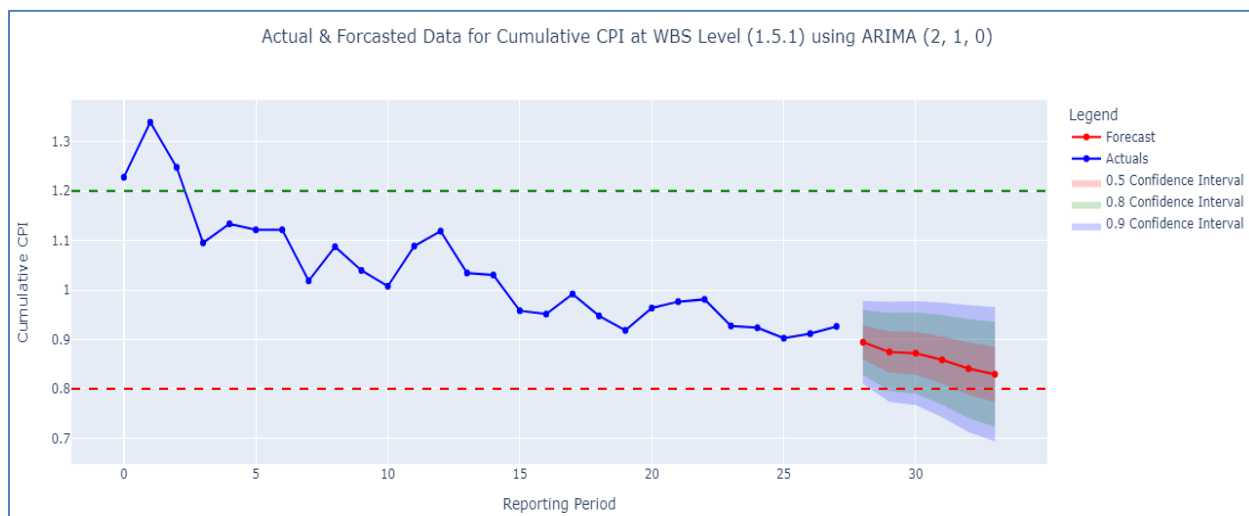


Figure 6. Cumulative SPI model, for WBS Element 1.5.1

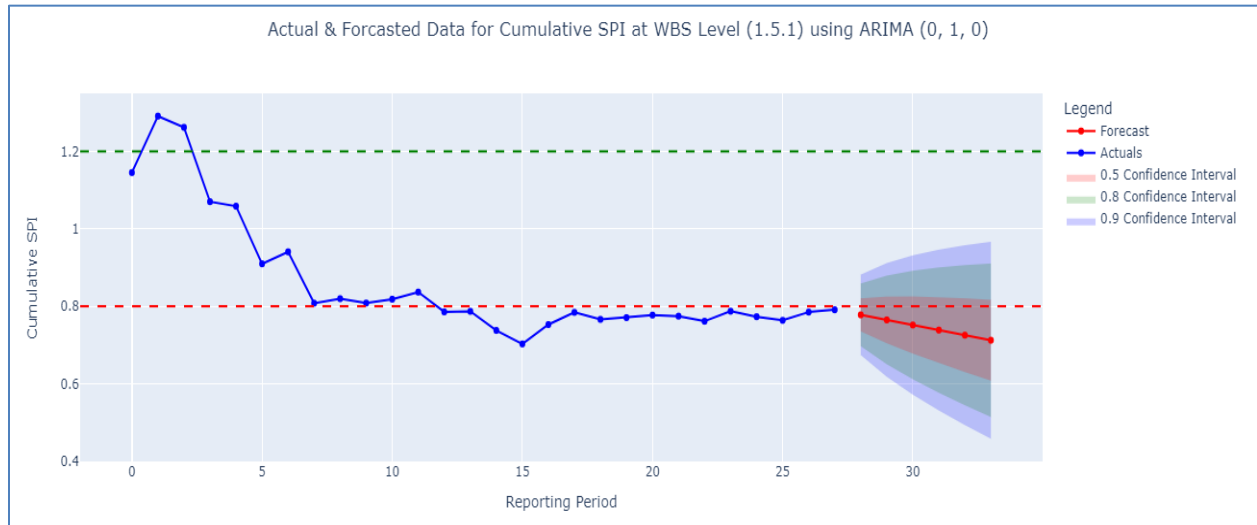


Figure 7. Cumulative CPI model, for WBS Element 1.5.1.1

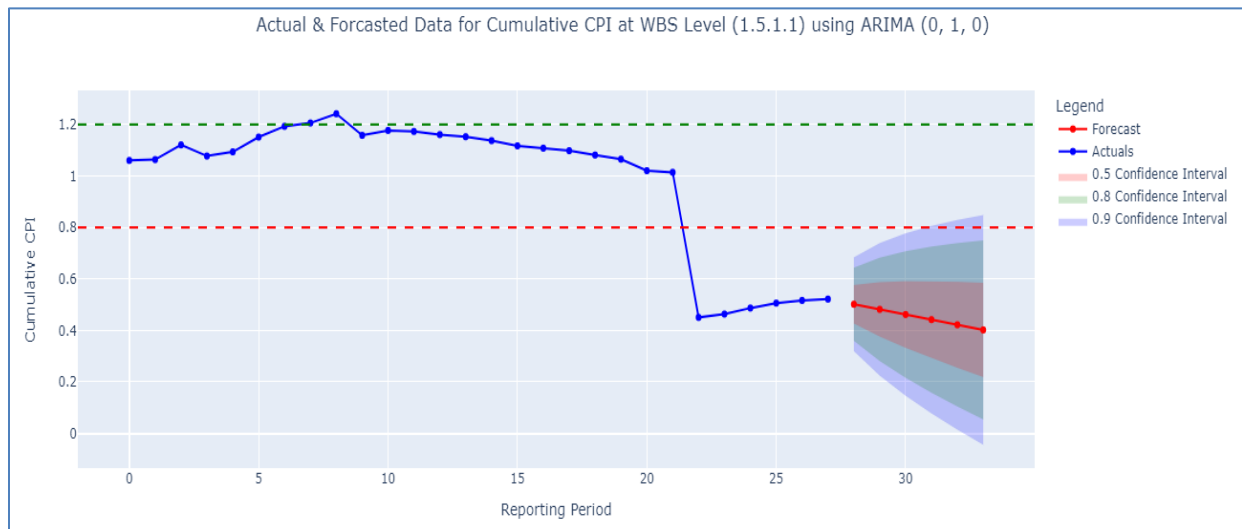
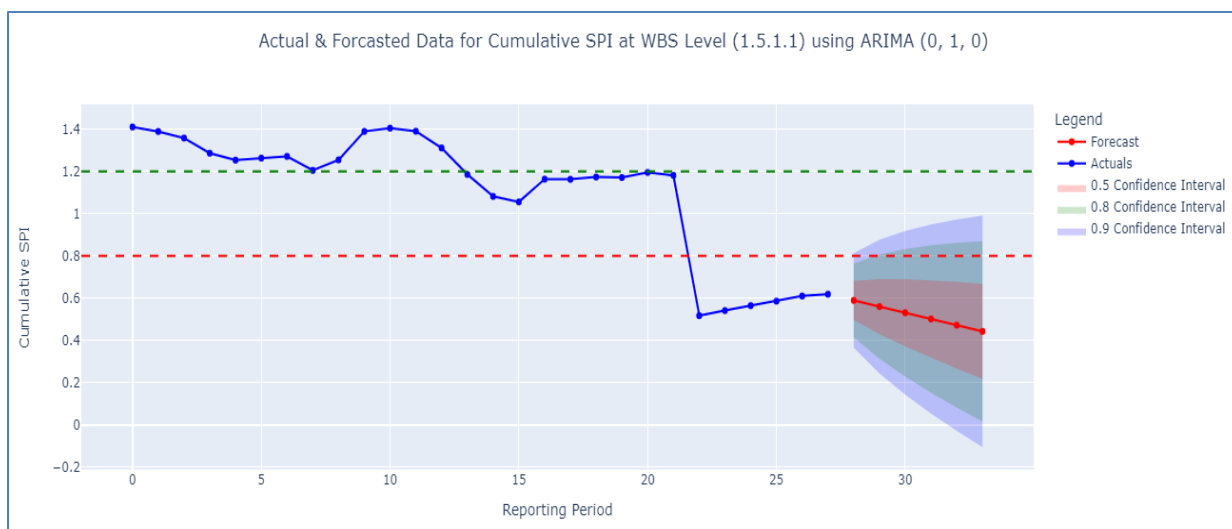


Figure 8. Cumulative SPI model, for WBS Element 1.5.1.1



The notional results of the cumulative CPI and cumulative SPI model forecasts, for each WBS element, were also used to calculate the IEAC values of each WBS element shown in Table 3 and Table 4. Ideally the IEAC and BAC should be close to the same value. Significantly larger IEAC values compared to the BAC typically indicate contract cost or schedule growth. A large, positive percent difference would indicate this occurred in the results. As shown in our notional example, the comparative values are highly variable, which more accurately reflects reality. At the third level of the WBS, the test and measurement equipment element, in particular, possesses an IEAC that is much greater than the planned BAC. Using this predictive IEAC metric may help acquisition analysts to periodically check the potential effect the cumulative CPI and cumulative SPI forecasts have on the program’s overall budget.

Table 3. Comparing Descriptive IEAC and BAC for Third Level WBS Elements

Third Level WBS Elements				
WBS Element	Name	IEAC	BAC	% Difference
1.5.1	Development Test & Evaluation	\$170,183	\$240,910	-29%
1.5.5	Test & Evaluation Support	\$968,086	\$497,392	95%
1.9.1	Test & Measurement Equipment	\$170,801	\$86,010	99%
1.9.2	Support & Handling Equipment	\$106,129	\$88,494	20%
1.10.2	Contractor Technical Support	\$245,399	\$230,504	6%

Table 4. Comparing Descriptive IEAC and BAC for Fourth Level WBS Elements

Fourth Level WBS Elements				
WBS Element	Name	IEAC	BAC	% Difference
1.1.1.2	Hull/Frame/Body/Cab	\$64,172	\$27,094	137%
1.1.1.6	Vehicle Electronics	\$227,919	\$180,301	26%
1.1.1.13	Special Equipment	\$513,538	\$318,294	61%
1.8.1.3	Test & Measurement Equipment (Electronics/Avionics)	\$200,637	\$105,109	91%
1.9.2.1	Support & Handling Equipment (Airframe/Hull/Vehicle)	\$92,070	\$16,010	475%

4. Macroeconomic Regression

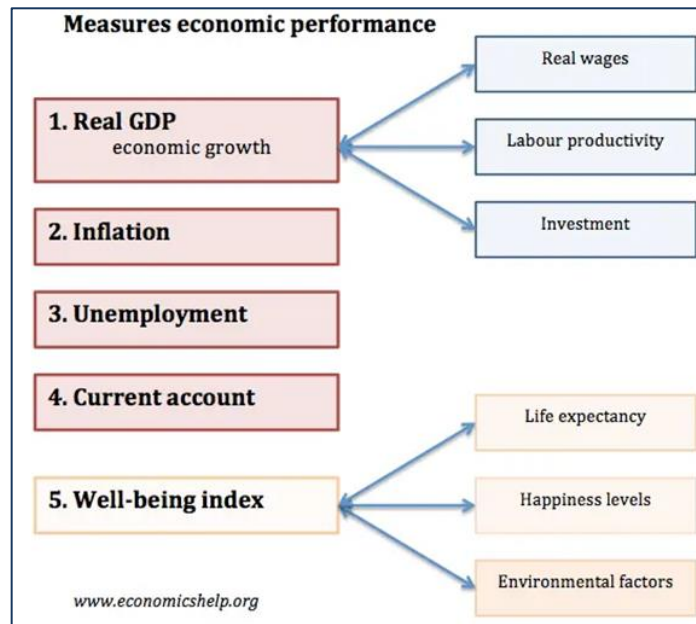
DoD acquisition program issues are not limited to activities and events that are internal to the program. External events may also drive cost and schedule deviations from a planned baseline. In discussing the root causes of Nunn-McCurdy acquisition breaches, Blickstein et al. (2012) find that exogenous events, such as industrial base changes, can lead to critical cost and schedule growth. As an example, evolution in the commercial satellite industry is cited as a root cause for a breach in the Wideband Global Satellite (WGS) program (Blickstein et al., 2012). Arena et al. (2014) assert that external considerations can impact program stability, citing overall demand for production function inputs and technology improvement as specific factors. The authors conclude that exogenous events necessitate greater monitoring in order to avoid or mitigate significant acquisition issues (Arena et al., 2014).

Economic indicators were selected as independent variables to evaluate against acquisition performance. Inflation and unemployment are two predominant measures of interest because strong anecdotal evidence exists for their influence on firm production in recent years. As one example, recent levels of inflation led to acquisition instability that manifested as issues in talent retention for government contractors, especially small firms, that ultimately threatened the defense industrial base (Overman 2022). Moreover, the U.S. economy has experienced exceptionally tight labor markets since 2020, which has led to labor recruitment challenges (Waddell and Macaluso, 2022). DoD system contractors attempting to hire staff for a new development or production contract may encounter staffing shortfalls as a result.

Empirical evidence also exists for how inflation and unemployment are measures of economic performance (Figure 9). Okun's law, for instance, highlights the importance of unemployment to

measuring overall economic performance, like gross output. The law measures the negative linear association between cyclical unemployment and the output gap (Equation 6), where cyclical unemployment is equal to the actual rate of unemployment minus the natural rate of unemployment and the output gap is equal to the actual output minus the potential output. Okun’s Law implies that changes in unemployment will behave countercyclically with economic output growth, such that a decline in cyclical unemployment increases the output gap. In terms of inflation, higher inflation levels can have significant costs, such as random redistribution of wealth, relative price variations, and more uncertainty about economic planning that leads to fewer long run contracts (Ball and Cecchetti, 1990). DoD contractors may find that certain portions of the supply chain become relatively more expensive, which can impact contract performance. In addition, the greater uncertainty in contracting from higher inflation can lead to a lower probability that the integrated baseline accurately reflects emergent economic realities.

Figure 9. Measures of Economic Performance



$$\Delta \text{output gap} = \alpha + \beta(\Delta[\text{actual unemp rate} - \text{natural unemp rate}]) + \epsilon$$

Equation 6. Okun’s Law

Therefore, the study team statistically examined whether these factors influence acquisition program health in terms of cost and schedule performance. The following research question was posed to guide the analysis. Do economic indicators that have a significant relationship with gross output, possess comparable relationships with DoD acquisition performance?

Economic data was collected from the Bureau of Labor Statistics (BLS). Specifically, the producer price index (PPI), which measures the change in domestic producers selling prices for their output, was used as a measure for inflation. Likewise, the study team collected local area

unemployment statistics (LAUS) to serve as measurements for inflation. Several normalization procedures are executed to integrate the macroeconomic data with the IPMDAR data sets and to conduct regression-based analysis at varying levels, to include a producer's industry and production locations.

The IPMDAR file format specification includes fields for contractor information by WBS element. The study team collated the contractor data, such as contractor and sub-contractor name and ID. Next, the team used the Defense Logistics Agency (DLA) to identify commercial and government entity (CAGE) codes and unique entity identifiers (UEIs) for each of contractors listed in the IPMDARs. A contractor table is subsequently produced, which includes contractor location information at the county level (e.g., Federal Information Processing Standards [FIPS] codes) based on the CAGE codes and UEIs and the firm's primary North American Industry Classification System (NAICS) codes. The NAICS code lists the specific industry that a firm is associated with. Ultimately, the completed contractor table includes foreign keys, such as FIPS and NAICS code that allows the analyst to interface with both the BLS macroeconomic data and the IPMDAR cost and schedule performance tables.

Once the normalization is completed, regressions are executed with CPI and SPI as the variables of interest and unemployment and inflation as the independent variables. Inflation regressions may be executed nationally and by NAICS. Unemployment regressions are executed by FIPS codes at the national, state, or county level or by NAICS codes. Both unemployment and inflation regressions are run at varying lagged rates, to include zero lag, under the assumption that some amount of time may be necessary for the macroeconomic parameters to influence contract performance. Measures of contract performance, like cumulative CPI and cumulative SPI, are calculated after rolling up the raw contract and schedule data by the level of interest (e.g., work location or industry). An example for unemployment as a single independent regressor with an intercept and error term is provided below (Equation 7).

$$CPI_{cumulative} = \beta_{Intercept} + \beta_1 Unemployment_{rate}_{lag=t} + \varepsilon$$

Equation 7. Unemployment Regression

Results have been executed for several programs to date. Notional results are displayed in Figure 10, showing a positive and statistically significant relationship between 12 month lagged unemployment at the county level for Dallas County, TX and the cumulative contract CPI. One potential interpretation is that a "tight" labor market, exacerbated by limited labor supply with the requisite human capital to manufacture DoD systems, leads to worse acquisition performance. A tight labor market occurs when the actual unemployment rate is less than the natural rate of ~4.75%, leading to a negative cyclical rate. The tight labor market portion of this graph, where the unemployment is less than ~4.75% aligns with Okun's law. On the contrary, the positive relationship of contract performance and higher levels of unemployment provides an inverse relationship to Okun's law. Reason exists for the results to diverge, as macroeconomic performance and contract performance are unique measures. The continuous positive

relationship may occur because higher levels of unemployment entail lower competition for labor supply, which enables defense contractors to hire and appropriately staff their contracts.

Figure 10. Unemployment and Cumulative CPI for Dallas County, TX



Overall, the specification for IPMDAR affords analysts an opportunity to explore how exogenous program factors may impact contract performance at relatively high fidelity. Given inflation projections provided by entities like the Congressional Budget Office (CBO), analysts can even forecast future CPI or SPI performance with a prediction interval that is based on a regression's results. Contractor information is provided by WBS elements and macroeconomic information from organizations like the BLS, Bureau of Economic Analysis (BEA), and Census Bureau, is mature, stable, and very detailed. Thus, the analyst may examine how external program factors impact their programs of interest broadly, at key geographic areas of design and manufacturing, and for specific industrial sectors. The analytic value is potentially substantial given the current industrial base and supply chain considerations for DoD acquisition.

5. Path Forward

The study team explored the application of predictive acquisition analysis using a variety of time series analyses and regressions with lagged economic performance indicators. However, many additional techniques are available to test, in large part due to proliferation of applicable data, metrics, and data science tools. The continued expansion of the portfolio of IPMDAR data through direct support of MDAPs and coordination with OSD would provide additional examples and use cases to build and refine the existing methodological approaches upon.

The integration and exploration of additional economic factors is a second advantageous course of action. The Census Bureau, Federal Reserve, and National Bureau of Economic Analysis (NBER) are three additional data sources offering cross-sectional and longitudinal statistics on a variety of economic factors that may exogenously influence DoD acquisition performance. Future efforts may test and train the time series and macroeconomic models on both a broader set

of economic data and a portfolio of IPMDARs. This exploration of economic factors would help identify and verify best fitting analytics, and establish classifiers (i.e., objective and threshold values) to help inform decision makers when opportunities or risks are imminent so that preventive action can be taken.

Another analysis path could be evaluating additional data science and statistical techniques such as panel regressions or clustering and classification techniques. Panel regressions, which combine cross-sectional and time-series data to control for unobserved dependencies and endogeneity could be used for a variety of research purposes. In regards to panel regressions, future research efforts may identify if work performance in a specific region of the U.S. is impacting one, or many, projects over time. Moreover, this type of analysis may compare acquisition performance by service, command, program size, prime system integrator, etc. A second analysis path available is the application of clustering and classification techniques such as K-nearest neighbors (K-nn) or logistic regression. K-nn is a supervised classifier that minimizes distance of items from a centroid of measurements which, can be used to classify the completion risk of activities over time, based on variables of interest (e.g. budgeted work remaining, active months). A final analysis path is logistic regression, which is a probabilistic model of an event (e.g. acquisition issue or opportunity) taking place which can be used to correlate cost and schedule variables of interest (e.g. average CPI or task mid-point SPI) with total cost outcomes (e.g. over budget or under budget) at varying WBS levels.

6. Conclusions

The study was constrained by several limitations, the most notable was data accessibility, as the analysis was executed on a single, on-going program. Using data from an on-going program rather than using data from completed projects limits the ability to assess the overall level of improvement provided by time series analysis. Nonetheless, forecasts can be timely in alerting the program of directional trends in cost and schedule health relative to standard EVM. Adequate sample sizes are another limitation. In order to conduct ARIMA analysis, multiple years of cost and schedule data (without re-baselines) are necessary to create a properly sized consecutive time series dataset.

Nevertheless, the methodological approaches discussed may offer several advantages over current EVM practices. Time series forecasts on standard EVM data and the SEMs may be more responsive to trends in project performance relative to descriptive statistics, which may subsequently allow the analyst more time to identify and decompose potential cost and schedule risks, before they manifest into significant issues. While traditional time series analysis has historically been conducted on a portfolio of completed programs or projects, this framework is also set up to train, test, and execute analysis in real time on current projects. Moreover, conducting time series analysis in real time and at varying WBS levels may allow for improved root cause analysis. This application creates an opportunity to prevent potential cost and schedule issues from emerging. Finally, this analysis accounts for trends in external program factors, which can be a key reason for a cost or schedule breach that may impact program performance.

References

- Alexander, Sean. "Earned Value Management Systems (EVMS) Basic Concepts", Project Management Institute Washington, DC Chapter, 2002.
<https://apps.dtic.mil/sti/tr/pdf/ADA403490.pdf>
- Arena, Mark V., et al., *Management Perspectives Pertaining to Root Cause Analyses of Nunn-McCurdy Breaches: Contractor Motivations and Anticipating Breaches*, Vol. 6, Santa Monica, Calif.: RAND Corporation. MG-1171/6, 2014. As of January 27, 2019: <https://www.rand.org/pubs/monographs/MG1171z6.html>, 2014.
- Anton, Philip S., William Shelton, James Ryseff, Samantha Cohen, Grant Johnson, Stephen B. Joplin, David Kravitz, Megan McKernan, and Alejandro Vigo Camargo. "Early Predictive Indicators of Contractor Performance: A Data-Analytic Approach." RAND Corporation, 15-42, 2022.
- Ball, Laurence, Stephen G. Cecchetti, and Robert J. Gordon. "Inflation and uncertainty at short and long horizons." *Brookings Papers on Economic Activity* 1990.1 (1990): 215-254.
- Blickstein, Irv, et al. "Methodologies in analyzing the root causes of Nunn-McCurdy breaches." (2012).
- Cryer, Jonathan D., and Kung-Sik Chan. Time series regression models. *Time series analysis: with applications in R*: 249-276, 2008.
- Data Item Description (DID) for Integrated Program Management Data and Analysis Report (IPMDAR), Number: DI-MGMT-81861C, Approval Date: 20210830.
- Department of Defense, Distribution Statements on DoD Technical Information., DoD Instruction 5230.24, Office of the Under Secretary of Defense for Research and Engineering, January 10, 2023.
- Department of Defense, Department of Defense Standard Practice Work Breakdown Structures for Defense Materiel Items, MIL-STD-881F, May 13, 2022.
- Dowling, Austin W. *Using Predictive Analytics to Detect Major Problems in Department of Defense Acquisition Programs*, Department of the Air Force Air University, Air Force Institute of Technology, March 2012.
- Integrated Program Management Data Analysis Report (IPMDAR) Implementation & Tailoring Guide, Department of Defense (DoD), 24 August 2021.
- IPMDAR Contract Performance Dataset, Version 1.0, File Format Specification, March 12, 2020
- IPMDAR Schedule Performance Dataset, Version 1.0, File Format Specification, March 12, 2020
- Keaton, C. G., White, E. D., & Unger, E. J., "Using earned value data to detect potential problems in acquisition contracts". *Journal of Cost Analysis and Parametrics*, 4(2), 148-159, 2011.
- Kwiatkowski, Denis, et al. "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?." *Journal of econometrics* 54.1-3 (1992): 159-178, 1992.

- Meier, Steven R. "Causal inferences on the cost overruns and schedule delays of large-scale US federal defense and intelligence acquisition programs." *Project Management Journal* 41.1 (2010): 28-39.
- Montgomery, Douglas C., Cheryl L. Jennings, and Murat Kulahci. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.
- National Reconnaissance Office (NRO), "Advanced Schedule Analysis Techniques", PowerPoint presentation, May 2021.
- National Reconnaissance Office (NRO), "Predictive Analytics: Schedule Execution Metrics", PowerPoint presentation, September 2022.
<https://ndiastorage.blob.core.usgovcloudapi.net/ndia/2022/ipmdFall/Bembers.pdf>
- Overman, T. *Defense authorization act provides inflationary relief for government contractors*. Reuters. December 20, 2022. <https://www.reuters.com/legal/legalindustry/defense-authorization-act-provides-inflationary-relief-government-contractors-2022-12-20/>
- Python API reference for plotly. <https://plotly.com/python-api-reference/>. Accessed May 2023.
- Schultheis, Candice, Ryan Carey, Robert Dunn. "An Analysis of Schedule Execution Metrics (SEMs) Developed by the National Reconnaissance Office (NRO)", Naval Post Graduate School Thesis, March 2021.
- Waddell, S. R., & Macaluso, C. (2022, September). *Changing recruiting practices and methods in the tight labor market*. Federal Reserve Bank of Richmond.
https://www.richmondfed.org/publications/research/economic_brief/2022/eb_22-36

