Journal of Cost Analysis and Parametrics, 4:148–159, 2011
Copyright © SCEA & ISPA
ISSN: 1941-658X print
DOI: 10.1080/1941658X.2011.628594

ICEAA
www.iceaaonline.com

Using Earned Value Data to Detect Potential Problems in Acquisition Contracts

C. GRANT KEATON, EDWARD D. WHITE, and
ERIC J. UNGER

Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio

Government contractors report earned value information to government agencies in monthly contract performance reports. Though major differences may exist in the data between subsequent contract performance reports, we know of no government effort to detect these occurrences. The identification of major changes may locate and isolate problems and, thus, prevent million and billion dollar cost and schedule overruns. In this study, we illustrate a proof of concept approach to identify changes in the cost performance index and the schedule performance index that may indicate problems with contract performance. We find the intuitive detection algorithm identifies changes in the cost performance index and the schedule performance index that correspond to large changes in the Estimate at Complete from 1 to 12 months out. The ability to detect unusual changes provides decision-makers with warnings of potential problems for acquisition contracts.

Background

Strains on the discretionary budget force military services to monitor cost and schedule performance for material acquisition closely. However, the deterioration of skills and personnel in the defense acquisition workforce decreases the Department of Defense’s (DoD) ability to provide adequate financial discipline (Morin, 2010). While the DoD presently addresses the reconstitution of the defense acquisition workforce, current acquisition analysts continue to manage an increasing workload. These analysts require new approaches to improve financial discipline in defense acquisition.

Several methods exist that may improve acquisition analysts’ ability to monitor cost and schedule performance. Specifically, analysts may develop more accurate estimate at complete (EAC) models and scrutinize changes in cost and schedule performance indices (Christensen, Antolini, & McKinney, 1995). The *Guide to Analysis of Contractor Cost Data* also provides guidance to acquisition analysts on the analysis of DoD contractor cost and schedule data (Headquarters Air Force Material Command, Financial Management, 1994). The intent of the guide is to aid acquisition programs in the reduction of cost growth and the improvement of visibility. The guide discusses numerous analytical techniques that focus on cost, schedule, and technical performance. Acquisition analysts study many of these measures (e.g., Cost Performance Index (CPI) and Schedule Performance Index (SPI)), in-practice.

Additionally, the guide offers direction on the use of problem analysis techniques. Problem analysis techniques include measures of cost and schedule efficiency, variance verification, management reserve analysis, manpower loading trend analysis, performance

Address correspondence to C. Grant Keaton, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson Air Force Base, OH 45433. E-mail: grantk@gmail.com

trends, forecasting by EAC function, and Over Target Baseline analysis. One particularly useful metric is the percent complete versus percent spent chart, which shows the cost and schedule performance expectation in the form of the CPI. A CPI slope of less than 1 indicates a program is spending more than expected. These deviations from the normal percent complete-percent spent line may indicate cost problems for an acquisition program. Similarly, deviations in the percent complete vs. percent scheduled chart in the form of the SPI may show schedule problems in an acquisition contract. A SPI slope of less than 1 indicates a program is behind in anticipated schedule.

With these tracking methods in mind, this research highlights to program analysts and DoD leadership a proof of concept approach for identifying problems within acquisition contracts in real-time using earned value data. Particularly, we center our research on the following: can we detect changes in acquisition contracts with an algorithm given at least the first three months of CPI and SPI data? If we can detect a change, how long does a change exist before we successfully identify it? We illustrate and test the ability of a forecasting algorithm to detect statistically significant changes in acquisition contracts' CPI and SPI. These detections identify contract areas that face or are at risk of ongoing cost overruns and schedule delays. Although program managers can use this information to aid analysis and exploration of program issues, this approach is not a substitute for expertise and understandings of their respective programs.

Researchers apply change detection to identify when system characteristics change. Even though multiple definitions and interpretations of "change" exist depending on the context or field of study, typically definitions of change detection focus on time-dependency. Specifically, abruptness, not necessarily magnitude, characterizes system change (Basseville & Nikiforov, 1993). Signal processing (Borodkin & Mottl', 1976; Cohen, 1987), time series analysis (Box, Jenkins, & Reinsel, 1994; Dasgupta & Forrest, 1996; Makridakis, Wheelwright, & Hyndman, 1998), automatic control (Willsky, 1976), and industrial quality control (Shewhart, 1931; Woodward & Goldsmith, 1964; Duncan, 1986) are some fields that apply change detection techniques. However, increases in information availability and advances in computer processing power provide new opportunities for change detection research (Cios & Moore, 2002; Venkatesh, 2007.) In this analysis, we employ autoregressive integrated moving average (ARIMA) models to study change detection. Because we only discuss those models investigated, we do not cover the entire spectrum of ARIMA models. We direct the reader to Box, Jenkins, and Reinsel (1994) or Makridakis, Wheelwright, and Hyndman (1998) for a complete discussion of time series analysis.

Database

The Defense Cost and Resource Center (DCARC) hosts a major collection of detailed earned value (EV) data for DoD acquisition contracts. These data include monthly contract performance reports (CPR), contract history files, and other EV and programmatic data submissions directly from program offices. For this analysis, we use EV history files available in DCARC.

Contract history files contain multiple entries for earned value information by month. Specifically for our analysis, the contract history files include the actual cost of work performed (ACWP), budgeted cost of work performed (BCWP), and the budgeted cost of work scheduled (BCWS). Because DoD and the American National Standards Institute maintain specific requirements and instructions for these measures, we assume the data provide a framework for reliable measurement (OUSD(AT&L)ARA/AM(SO), 2005; NDIA/PMSC, 2009). We limit our analysis database to history files for research, development, test, and evaluation (RDT&E) contracts in DCARC. We limited the database to RDT&E contracts because of their relatively high risk relative to production contracts, increasing the possibility of validating our approach. That is, if our proposed method does not work in a riskier spectrum, it's unlikely to work on less risky contracts.

In an internal query of DCARC, we identify 813 that meet our database specifications. Of these 813 contracts, we could only locate 787 files in the actual database itself. The different file types of the search results (e.g., .pdf and .trn) reduce the number of files we can download into a database from 787 to 165 because we cannot extract all the data automatically (without an extensive manual data entry effort, which we comment on later in the results section). Lastly, of the 165 files we can access, we discover 32 unique contract history files for RDT&E contracts. We eliminate one history file due to data inconsistencies. We list in Tables 1 and 2 the number of contracts in the research database by military handbook type and military service, respectively. We do not impose a contract start date or end date constraint on the research database due to the small number of history files we gather from DCARC; however, the start date for all but one contract is after January 1, 2000 (see Table 3).

As with any database, there are limitations with respect to the data source. Acquisition contract history files offer some benefits, but pose many obstacles to analysis. These

TABLE 1 Number of contracts by military handbook type

Military handbook type	Number of contracts
Aircraft	8
Electronic/Automated software	13
Missile	3
Ship	1
Space	3
Surface vehicles	2
System of systems	1
Total	31

TABLE 2 Number of contracts by military service

Military service	Number of contracts
Air Force	11
Army	7
Navy	12
Department of Defense	1
Total	31

TABLE 3 Number of contracts by contract start date

Contract start date	Number of contracts
1 Jan 1995–31 Dec 1999	1
1 Jan 2000–31 Dec 2004	11
1 Jan 2005–31 Dec 2009	19
Total	31

TABLE 4 Descriptive statistics for the percent data coverage for all 31 contracts in the earned value modeling database (Rosado, 2011)

Minimum contract percent coverage	12%
Maximum contract percent coverage	100%
Median contract percent coverage	48%
Mean contract percent coverage	51%
Standard deviation	25%

obstacles are four-fold. First, a contract history file is effectively a concatenation of sequential monthly CPRs. Often monthly CPRs contain inaccuracies, which program offices work with the contractor to correct. CPR re-submissions to DCARC are evidence of this issue. However, in some instances systematic errors persist in the contract history files we collect. We attempt to resolve these data issues with the appropriate monthly CPRs or the applicable CPR resubmissions.

Second, a contract history file does not always contain the full time series. One reason for partial time series is that many program offices update their contract history files on an annual basis. Thus, a researcher who collects history files between updates may not acquire the additions to the time series since the last release. In Table 4, we list the descriptive statistics of the percent data coverage for all 31 contracts in the modeling database. We calculate percent coverage by comparing the contract start date and contract end date to the available months of data in the contract history files. Generally, the length of the time series in a contract history is shorter than the time from contract start date to present. Thus, some of the contract history files we use have fewer months than the contract's actual number of months to-date.

Third, the flexibility in electronic submission formats permitted by the CPR-governing data item description creates data accessibility issues for cross-program analysis that individual program offices may not face (OUSD(AT&L)ARA/AM(SO), 2005). Specifically, our data processing and management resources cannot process all file types that contractors submit. Individual program offices likely do not have this issue because they have a direct relationship with the contractor and can specify an electronic format both can handle easily.

A final limitation on our research database is the restriction on the file size program offices can upload to DCARC. That is, file sizes that are too large to submit are unavailable in DCARC and, thus, impact the number of contract history files we collect. As a result, DCARC inadvertently filters available contract history files.

Methodology

We construct our research database with entries for ACWP, BCWP, and BCWS with respect to report date for each contract history file. We sort these using the Work Breakdown Structure (WBS) level as the criterion. For the WBS level criterion, we sort the data by level 1 and sum the values within the level. These sums are cumulative values for ACWP, BCWP, and BCWS. We limit the sort criteria to WBS level 1, but conceivably can use levels 2 and 3 also. Data for WBS levels greater than 3 are problematic because fewer contracts report at each lower level and, thus, reduce the sample size increasingly. Different sample sizes create data comparison issues between acquisition contracts.

We compute monthly ACWP, BCWP, and BCWS values and monthly and cumulative analytic earned value measures for the level 1 data. We calculate the analytic EV measures of monthly and cumulative CPI and the monthly and cumulative SPI. Because differences in the size (e.g., budget at complete), contract length, and inflation can complicate comparisons among contracts, we need to address how we deal with these issues.

First, the importance of a change in ACWP, BCWP, or BCWS is relative to the size of the particular contract. Although a change may be large in amount, the relative change may be small compared to the size of the overall contract. However, calculations for CPI and SPI control for contract size because changes in ACWP, BCWP, and BCWS are relative to one another. Therefore, this ratio for each program is already normalized.

Next, the length of a contract may influence how abruptly a change appears over an entire contract. Traditionally, EV analysts use a percent complete calculation to manage this concern. In this analysis, we focus on monthly changes, not changes throughout entire contracts. Therefore, contract length does not affect our analysis.

Lastly, the effect of inflation creates disparities in the value of money across time. We use 2010 as a base year (BY10$) to standardize costs in time. We gather the conversion rates from the 2010 release of Deputy Assistant Secretary of the Air Force for Cost and Economics (SAF/FMC) inflation tables (SAF/FMC, 2010).

With the dataset established, we turn to time series to analyze our data. As mentioned in the previous section, ARIMA forecasting offers a logical approach to online change detection in earned value data. We theorize that patterns in cumulative ACWP, cumulative BCWP, and cumulative BCWS time series are distinguishable from data noise. We model these patterns to determine how we can best show real-time changes in the CPI and the SPI. Although we lack a large amount of data for any single program, our database has enough observations to confirm trends for several programs. Lastly, we expect historical cost and schedule performances to continue in the future.

We analyze our time series in JMP® version 9 (JMP®, 2010). The time series capability in JMP® includes ARIMA models, which we use to forecast EV data. The parameter test statistics and rank criteria we obtain from JMP® help us appraise each acquisition contract model in our research database. We record consistent time series characteristics to consider during model selection. Largely, we conduct our analysis using the Box-Jenkins approach (Box, Jenkins, & Reinsel, 1994) using autocorrelation functions (ACF), partial autocorrelation functions (PACF), and differencing functions. We plot each time series to examine if the means and variances are stationary for the ACWP, BCWP, and BCWS time series and employ the Augmented Dickey-Fuller test (ADF) to statistically test this at the 0.05 level of significance.

The ACF and PACF plots reveal potential autoregressive models, moving average models, or integrated models (differencing). We do not observe any statistical seasonal patterns. Therefore, we confine our model selection to non-seasonal ARIMA models that account for these characteristics. We use the ARIMA(p, d, q) model group function in JMP® to test models that meet the inclusive range of specifications for p, d, and q, where these parameters can either equal 0 or 1. These 2^3 combinations identify the eight potential time series models for modeling consideration.

The ARIMA model group function ranks models by the Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (SBC). The smaller the AIC and SBC values, the better rank the model earns (Akaike, 1974; Schwarz, 1978). The rank structure provided by the ARIMA model group function was consistent between AIC and SBC measures. These metrics for potential models distinguish ARIMA models ARI(1,1), IMA(1,1), I(1), and ARIMA(1,1,1) from AR(1), MA(1), ARIMA(0,0,0), and ARMA(1,1). Table 5 lists the number of contracts in which each model occurred in the top four ranks according to the AIC and SBC measures. Because the first four models listed appear in the top four model ranks for nearly every program, we choose to examine these models further.

TABLE 5 Top four ARIMA model occurrences as determined by the Akaike Information Criterion and Schwarz Bayesian Criterion

	Contracts		
ARIMA model	ACWP	BCWP	BCWS
ARI(1,1)	31	31	31
IMA(1,1)	30	30	30
I(1)	30	30	30
ARIMA(1,1,1)	30	30	30
ARIMA(0,0,0)	2	2	2
AR(1)	1	1	1
MA(1)	0	0	0
ARMA(1,1)	0	0	0

We validate the appropriateness of the high-occurrence model group [ARI(1,1), IMA(1,1), I(1), and ARIMA(1,1,1)] with tests of statistical significance for the terms in each model. Table 6 lists the number of contracts in which all parameters for a given model are statistically significant (for $\alpha = 0.05$). We find that three out of four models in the high-occurrence group have one or more variables that are not statistically significant for approximately half of the contracts in the research database. From the results, it is very clear that the I(1) model performs well against the model rank criteria and passes the tests of statistical significance for nearly all contracts. For this reason, we discard the other time series models and adopt the I(1) model only for our change detection algorithm. Mathematically, the I(1) is expressed as $Y_t = Y_{t-1} + \varepsilon$, where Y_t is the current month's cumulative AWCP, BCWP, or BCWS and Y_{t-1} is the previous month's cumulative AWCP, BCWP, or BCWS. The random error term, ε, represents the stochastic variability in the model.

The choice of the ARIMA I(1) model implies that better statistical time series models for cumulative AWCP, BCWP, and BCWS require differencing the current cumulative values with the previous month's cumulative values. This differencing simplifies quite nicely to now just modeling the current monthly AWCP, BCWP, and BCWS. Therefore, our change detection algorithm uses statistical differences to monitor real-time changes in the monthly CPI and SPI observations. We theorize changes in the CPI, and the SPI may indicate contract problems because these measures are the slopes of the percent complete vs. percent spent and percent complete vs. percent scheduled plots, respectively. We define

TABLE 6 Number of contracts out of 31 in the modeling database that possess statistically significant parameters ($\alpha = 0.05$) for the various ARIMA Time Series Models

	Contracts		
ARIMA model	ACWP	BCWP	BCWS
I(1)	28	27	27
IMA(1,1)	16	11	9
ARI(1)	13	11	11
ARIMA(1,1,1)	10	9	10

a difference as a CPI or a SPI value statistically different from 1 (i.e., on time and on schedule).

We use Chebychev's Theorem (Newbold, Carlson, & Thorne, 2010) to establish a region whereby the algorithm flags any monthly CPI or SPI value outside this boundary. Given our small sample size, we chose to adopt a conservative statistical approach. Since Chebychev's Theorem requires no such assumptions as normality and/or constant variance to generate a region, we felt confident in this method. The specific boundary established for our change detection algorithm is $\bar{x} \pm ks$. Since ideally the CPI and SPI should be 1, we set $\bar{x} = 1$. We test the sensitivity of the algorithm for a series of standard deviations to trade off false detections (Type I errors) with missed detections (Type II errors); specifically, we vary k from 0.5 to 3. For an accurate estimate of the standard deviation (s), we do not begin change detection until the fourth monthly observation of the CPI and SPI. That is, the first observation for which we attempt to detect a change in each time series is the fourth month's observation.

When the algorithm detects a change of a contract's monthly CPI or SPI outside this established interval, we also look for major changes in the contractor's estimate at complete (EAC). We theorize months that indicate a detectable change in the monthly CPI or SPI will lead or correspond to major changes in the contractor EAC. A change in the contractor EAC is a significant event. Formally, the company under contract acknowledges that it likely cannot complete the work required at or within the dollar value of the current EAC. We define $\%\Delta\text{EAC}$ as:

$$\left(\frac{EAC_{Current} - EAC_{Previous}}{EAC_{Previous}}\right) * 100\%,$$

where current EAC refers to the most recent EAC, while previous EAC is the previous month's EAC. Lastly, we categorize $\%\Delta\text{EAC}$ into: $\%\Delta\text{EAC} \leq -10\%$, $-10\% < \%\Delta\text{EAC} \leq -5\%$, $5\% \leq \%\Delta\text{EAC} < 10\%$, and $10\% \leq \%\Delta\text{EAC}$. We choose these categories to characterize major EAC changes because changes within $+/-5\%$ occur frequently and, therefore, likely represent normal data noise. Changes of at least 5% appear much less frequently and, thus, we theorize are indicative of major performance changes.

Results and Discussion

After determining that the I(1) ARIMA model best modeled the contracts in our database with respect to cumulative ACWP, BCWP, and BCWS, we tested the algorithm. Overall, we found 99 months had major percentage changes in the EAC out of 1,094 potential months for all 31 contracts. Logically, the number of changes detected in the CPI and SPI increased with greater algorithm sensitivity, i.e., as we decreased the difference between the lower control limit (LCL) and upper control limit (UCL). [Note: The lower and upper part of the interval was formed from $1 \pm ks$.]

For perspective, the most sensitive algorithm we tested ($k = 0.5$) identified 550 and 549 changes in the CPI and SPI, respectively. This algorithm sensitivity detected changes in approximately half of the 1,094 observations in the research database and about five times the number of major EAC changes that occurred during the same month as the detections. The least sensitive algorithm ($k = 3.0$) detected statistical changes in the CPI and SPI for 89 and 75 observations, respectively. Therefore, the least sensitive algorithm we tested detected changes in less than 10% of observations and less than 80% of the number of major EAC changes that occurred during the same month as the detections.

For both CPI and SPI across the different standard deviations, observations exceeded the LCL more frequently than the UCL: 83% and 84%, respectively. The higher percentage of LCL detections does not imply that the algorithm is more sensitive to worsening cost and schedule performances. Rather, the algorithm detected a higher ratio of worsening than improving cost and schedule performances for the 31 contracts in our database. We interpret this as an indication that the I(1) model does a very good job of detecting degrading monthly CPI or SPI metrics. These detections alert analysts that further investigation is required.

We also investigated if a significant change in either the monthly CPI or SPI preceded a major change in the EAC. We determined that the algorithm did identify informational early detection relationships between CPI or SPI change and for all groups of major EAC changes. Changes in the CPI and SPI corresponded to major changes in the EAC as early as 12 months before the EAC change, with on average 4–5 months being the trigger point for likely detection. The percentage of detections grew as the time difference between the CPI or SPI detection decreased from the EAC change. Conversely, the number of non-detections decreased as time between detection and EAC change decreased. We display and illustrate these findings in Tables 7 and 8 and Figures 1 and 2. Although the appropriate upward and downward trends are evident in these figures, there are slight deviations from these

TABLE 7 Detection and non-detection percentages of varying EAC changes using significant change in CPI from 1 month to 12 months prior to the actual change in EAC for $\sigma = 0.5$

		Months before EAC change											
		12 (%)	11 (%)	10 (%)	9 (%)	8 (%)	7 (%)	6 (%)	5 (%)	4 (%)	3 (%)	2 (%)	1 (%)
Detect	$10 \le \%\Delta EAC$	28	32	36	33	37	39	37	43	47	51	47	51
	$5 \le \%\Delta EAC < 10$	41	38	32	38	38	44	44	47	56	62	59	59
	$-10 < \%\Delta EAC \le -5$	25	33	25	33	42	33	33	33	50	33	58	42
	$\%\Delta EAC \le -10$	38	46	46	54	54	54	54	54	62	62	69	69
Does not detect	$10 \le \%\Delta EAC$	72	68	64	67	63	61	63	57	53	49	53	49
	$5 \le \%\Delta EAC < 10$	59	62	68	62	62	56	56	53	44	38	41	41
	$-10 < \%\Delta EAC \le -5$	75	67	75	67	58	67	67	67	50	67	42	58
	$\%\Delta EAC \le -10$	62	54	54	46	46	46	46	46	38	38	31	31

TABLE 8 Detection and non-detection percentages of varying EAC changes using significant change in SPI from 1 month to 12 months prior to the actual change in EAC for $\sigma = 0.5$

		Months before EAC change											
		12 (%)	11 (%)	10 (%)	9 (%)	8 (%)	7 (%)	6 (%)	5 (%)	4 (%)	3 (%)	2 (%)	1 (%)
Detect	$10 \le \%\Delta EAC$	30	34	34	33	37	45	43	43	41	49	57	47
	$5 \le \%\Delta EAC < 10$	38	29	35	47	44	44	47	56	59	59	65	65
	$-10 < \%\Delta EAC \le -5$	25	33	25	42	33	42	42	50	50	50	67	75
	$\%\Delta EAC \le -10$	38	46	54	38	38	31	38	46	62	62	46	62
Does not detect	$10 \le \%\Delta EAC$	70	66	66	67	63	55	57	57	59	51	43	53
	$5 \le \%\Delta EAC < 10$	62	71	65	53	56	56	53	44	41	41	35	35
	$-10 < \%\Delta EAC \le -5$	75	67	75	58	67	58	58	50	50	50	33	25
	$\%\Delta EAC \le -10$	62	54	46	62	62	69	62	54	38	38	54	38

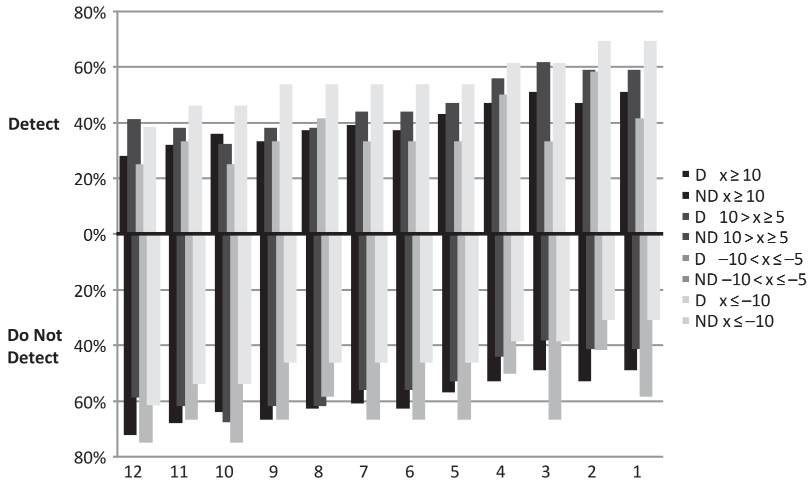

FIGURE 1 Graph of the detect/don't detect percentages as shown in Table 7. Horizontal axis represents the number of months prior to an EAC change (× in legend).

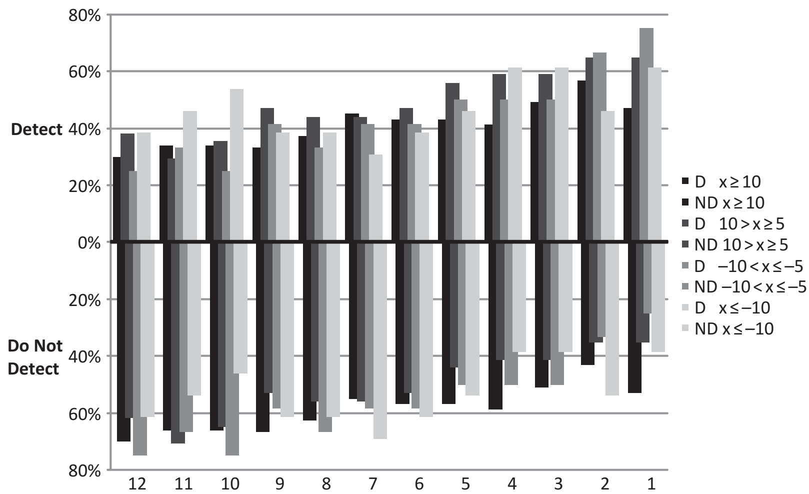

FIGURE 2 Graph of the detect/don't detect percentages as shown in Table 8. Horizontal axis represents the number of months prior to an EAC change (× in legend).

overall trends. We expected this given our small sample size of 31 contracts and it does not detract from the overall conclusion that the I(1) proof of concept appears capable of detecting future major EAC changes.

Next, we investigated whether or not simultaneous CPI and SPI changes resulted in a major change in the EAC for the same month; the algorithm identified 185 occurrences of simultaneous CPI and SPI changes. Of the 185 occurrences, 13 corresponded to major changes (as described earlier and denoted in Table 9) in the EAC. The algorithm detected all these major changes in the EAC (0% missed detection rate). Table 9 lists the numbers and percentages of detections by group of major EAC change. When there were simultaneous detections of the CPI and SPI change, over half (54%) of the contracts experienced at least a 10% increase in EACs.

Continuing, we examined the relationship between sequential detections for the CPI and SPI and a subsequent major change in the EAC. Specifically, we analyzed whether a

TABLE 9 Simultaneous CPI and SPI detections during the same month as major EAC change

	Same month detections	
% Change in EAC	# Detections	% Detections
$10 \leq \text{EAC}$	7	54
$5 \leq \text{EAC} < 10$	3	23
$-10 < \text{EAC} \leq -5$	2	15
$\text{EAC} \leq -10$	1	8
Total	13	100

detection in the CPI or the SPI was followed by a detection in the opposite index (CPI or SPI) during the next 12 months. If a sequential detection was identified, we looked for a major change in the EAC during the 12 months after the second detection; we found no such occurrences.

Lastly, given this 100% detection of simultaneous CPI and SPI change along with a major change in the EAC, we wanted to determine if this held true for contracts outside of the modeling database. As mentioned previously, we electronically collected all the contracts that we could from DCARC. For a modicum of validation to this detection rate, we randomly pulled five more contracts but this time manually entered all the earned value information. [Note: We found this to be very time consuming, and offer that standardizing DCARC to have the data stored in either Microsoft Excel or Access format would greatly aid analysts of this EV data.] We recognize that like 31, 5 is not a large number, however, we did notice that only twice did we detect simultaneous CPI and SPI change along with a major change in the EAC. And in both instances, the I(1) proof of concept model detected these for again a 100% detection rate.

Conclusions

Our proof of concept analysis of earned value data reveals we can detect changes in acquisition contract performance, namely the cumulative AWCP, BCWP, and BCWS. We determined that the best statistical ARIMA model consisted of a first differencing equation, resulting in using monthly CPI and SPI data. From this, we developed an algorithm using updating lower and upper control limits to detect these changes. We found that the change detection algorithm identifies worsening more often than improving cost and schedule performances.

Additionally, we found that the detections led major changes in the EAC by as much as 12 months, with 4–5 months appearing to be the trigger point. That is, this initial proof of concept algorithm descriptively appears to be able to equally detect a problem a month ahead at the same likelihood of success as 4 to 5 months. Future research is aimed at building upon this proof of concept and stretching out this time window to 6 or 12 months out. The percentage of detections for major EAC changes increases as the time between detection and EAC decreases. Lastly, approximately 77% of simultaneous changes detected for the CPI and SPI corresponded to large EAC increases. Sequential CPI-SPI detections did not yield any major future EAC changes.

One noteworthy issue we encountered during this analysis was what actually constitutes a problem in contract performance. We used EAC as a problem confirmation measure, but EAC as a problem indicator presented difficulty. The difficulty was EACs may increase because contracts run over cost or because the contract took on a larger scope and requirements. We differentiated between overrun increases from scope increases by categorizing EAC growth percentages given detection or no detection. If the algorithm did not detect a change in the CPI or SPI and a large percentage increase in EAC occurred, we assumed the increase in EAC was scope-related. We recognize that as a big assumption and again future research is aimed at bettering this determination.

The ability to detect problems in acquisition contracts offers DoD leadership a method to monitor cost and schedule performance in real-time. The benefit of real-time analysis in defense acquisition is two-fold. First, the identification of contracts that transform suddenly—and significantly—from good or normal performance to bad performance offers a great capability to program managers and DoD leadership. With real-time problem information, these leaders can identify, isolate, and potentially avoid major cost and schedule overruns. In the future, major cost and schedule overruns may pose serious concerns for acquisition contracts due to the likelihood of greater fiscal scrutiny.

Secondly, automated real-time analysis helps solve a principal concern of many acquisition leaders. Specifically, automated analysis alleviates some of the strains caused by low personnel levels in the acquisition workforce. To be clear, this does not remove the responsibility of potential users to understand the limitations of this algorithm and method. The algorithm and method provide a way to gain insight into an acquisition contract in addition to or in absence of other information and acquisition professionals.

Despite our limited sample size of 31 contracts, we are pleased with the proof of concept results achieved. We believe the ARIMA algorithm is both statistically sound, overwhelming the model of choice as indicated by our data, and easy to use. The algorithm only needs monthly CPI and SPI data. As with any other tool, the algorithm does not replace the person monitoring a program. Instead, it adds another tool in the toolbox to use when monitoring earned value data in real-time.

Acknowledgments

The authors appreciate the support of Mr. Steven Miller from the Office of the Secretary of Defense, Cost Assessment and Program Evaluation, in sponsoring and funding this research. His support does not constitute an endorsement of the authors' analysis. The authors also wish to thank the anonymous reviewers for their constructive comments that aided in improving this article.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716.

Basseville, M., & Nikiforov, I. V. (1993). *Detection of abrupt changes: Theory and application*. Englewood Cliffs, NJ: PTR Prentice Hall, Inc.

Borodkin, L., & Mottl', V. (1976). Algorithm for finding the jump times of random process equation parameters. *Automation and Remote Control*, *37*(6), 23–32.

Box, G. E., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control*. Upper Saddle River, NJ: Prentice Hall.

Christensen, D. S., Antolini, R. C., & McKinney, J. W. (1995). A review of estimate at completion research. *Journal of Cost Analysis and Management*, 41–62.

Cios, K. J., & Moore, G. W. (2002). Uniqueness of medical data mining. *Artificial Intelligence in Medicine*, *26*(1), 1–24.

Cohen, A. (1987). *Biomedical signal processing* (Vol. 1 and 2). Boca Raton, FL: CRC Press.

Dasgupta, D., & Forrest, S. (1996). Novelty detection in time series data using ideas from immunology. *5th International Conference on Intelligent Systems*. Reno, NV.

Duncan, A. (1986). *Quality control and industrial statistics*. Homewood, IL: Richard D. Irwin, Inc.

Headquarters Air Force Materiel Command, Financial Management. (1994). *Guide to Analysis of Contractor Cost Data*. Maxwell AFB, AL: AUL/LDEA.

JMP®. 2010. Version 9.0. Cary, NC: SAS Institute Inc.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and applications*. New York, NY: John Wiley & Sons Inc.

Morin, J. M. (2010). Achieving acquisition excellence in the Air Force: A financial management perspective. *Armed Forces Comptroller*, *55*(2), 8–12.

NDIA/PMSC. (2009). *Earned value management systems intent guide*. Arlington, VA: National Defense Industrial Association.

Newbold, P., Carlson, W. L., & Thorne, B. (2010). *Statistics for business and economics*. Upper Saddle River, NJ: Prentice Hall.

OUSD(AT&L)ARA/AM(SO). (2005). Data item description—Contract performance report (CPR). OUSD(AT&L).

Rosado, W. (2011). Comparison of Development Test and Evaluation and Overall Program Estimate at Completion. (Masters thesis). Air Force Institute of Technology, Wright-Patterson AFB, OH.

SAF/FMC. (2010, January 8). USAF inflation indices base year 2010.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464.

Shewhart, W. (1931). *Economic control of quality manufactured product*. Princeton, NJ: D. Van Nostrand.

Venkatesh, S. (2007). Recognition of emergent human behaviour in a smart home: A data mining approach. *Pervasive and Mobile Computing*, *3*(2), 95–116.

Willsky, A.S. (1976). A survey of design methods for failure detection in dynamic systems. *Automatica*, *12*, 601–611.

Woodward, R., & Goldsmith, P. (1964). *Cumulative sum techniques*. Edinburgh, UK: Oliver and Boyd.

About the Authors

First Lieutenant C. Grant Keaton is a space systems cost analyst at the Air Force Cost Analysis Agency. He earned a B.S. in Economics and a minor in Russian from the United States Air Force Academy and an M.S. in Cost Analysis from the Air Force Institute of Technology. His current research focuses on time series analysis and quantitative content analysis.

Dr. Edward D. White is an Associate Professor in the Department of Mathematics and Statistics. He has served as a member of the AFIT faculty since the summer of 1998. Dr. White received his B.S. in Mathematics from the University of Tampa, his M.A.S. in Applied Statistics from The Ohio State University, and his Ph.D. in Statistics from Texas A&M University. His work has been published in various journals such as the *Air Force Journal of Logistics, Journal of Cost Analysis and Management, Defense Acquisition Review Journal, Cost Engineering, Journal of Public Procurement,* and the *Journal of Cost Analysis and Parametrics*, where he serves as co-editor. His primary research interests include statistical modeling and simulation.

Lieutenant Colonel Eric J. Unger is the Director of the Cost Analysis Graduate Program at the Air Force Institute of Technology. He received a B.A. in Mathematics and Economics from Northwestern University, an M.S. in Acquisition Management from the Air Force Institute of Technology, and a Ph.D. in Policy Analysis from the Pardee RAND Graduate School. He served previously as the Chief of Cost, MILSATCOM, at Los Angeles Air Force Base. His research focuses on the policy impact of quantitative cost analysis.