Trust, but Verify: An Improved Estimating Technique Using the Integrated Master Schedule (IMS)

Eric Lofgren, Technomics, Inc.
International Cost Estimating and Analysis Association (ICEAA) conference
10-13 June 2014

> **Trust, but verify** is a form of advice given which recommends that while a source of information might be considered reliable, one should perform additional research to verify that such information is accurate, or trustworthy. The original Russian proverb is a short rhyme which states, Доверяй, но проверяй (doveryai, no proveryai).
> -Wikipedia

It has long been the wonder of management why the Integrated Master Schedule (IMS) fails to give advanced warning of impending schedule delays. An estimator may follow authoritative guidance in analysis of schedule health using key metrics, supposing that such checks authenticate schedule realism. Why, then, do practitioners find themselves caught off guard by slips when their IMS appears in good health? Answers to this question follow when one attempts to independently trace IMS development across submissions. By analyzing an IMS evolves over time, this paper will show that early performance to baseline is a good measure of realized slip and provides the framework for a new estimating technique.

The Integrated Master Schedule is more than a tool for managing the time phasing and resource allocation of project activities. For a decision-maker, the value of an IMS is its ability to evaluate risk associated with schedule and provide early warning for schedule slip. This information gives insight into one of the three categories for total contract performance risk, the other two being cost and technical risk. In general, realized risk in any one area can only be offset at the expense of performance in one of both of the others (see Figure 1 below).[1] For example, if the IMS predicts a significant schedule delay, the decision-maker may allocate additional resources to the problem or reduce systems capabilities in order to maintain the contractual delivery date. Therefore, accurate and timely information on schedule risk provides flexibility in project management.

[1] There is often a positive correlation amongst cost, schedule and technical risk. Sometimes managers can only mitigate poor performance.

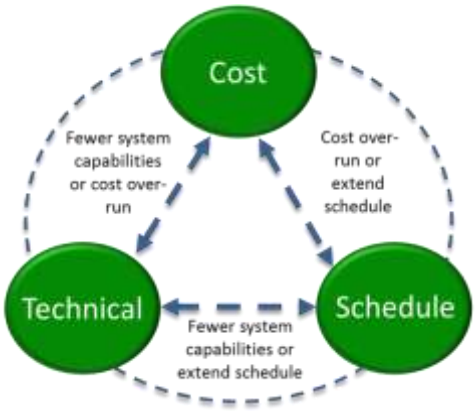


Figure 1: Cost, schedule and technical risk interrelationship

For many Major Defense Acquisition Program (MDAP) contracts, the IMS has a poor history of accurately reflecting schedule risk. In an analysis of 12 MDAP contracts,[2] schedule slips were not registered until late in the project. In Figure 2 below, one can see that the IMS largely does not indicate any delays until after half-way through the original schedule duration. In fact, as a project approaches its expected end date further delays are developed, creating a tail chase[3]. This situation gives decision-makers minimal leeway for management to make tradeoffs and implementing well-informed strategies.

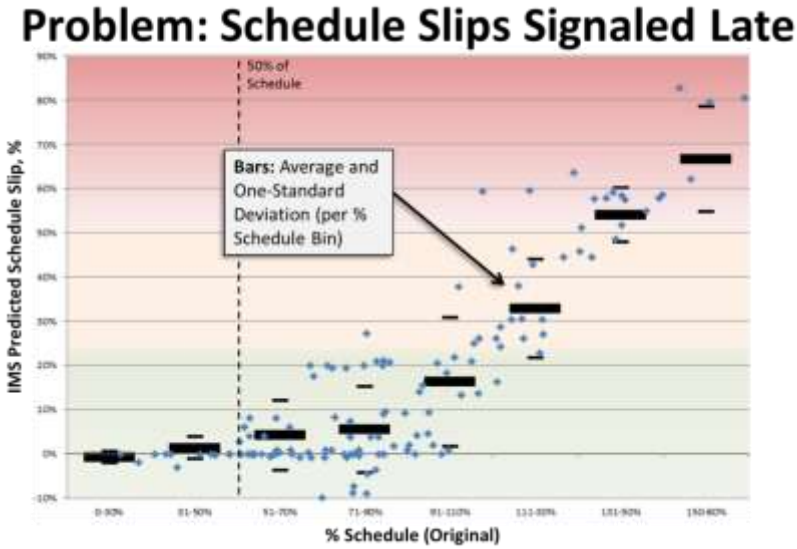


Figure 2: Predicted schedule slip reported by IMS submission

[2] MDAP contract data accessed through the Earned Value Management Central Repository (EVM-CR). This paper reflects an analysis of all contract schedules available. Data excluded: contracts without an approved Non-Disclosure Agreement (NDA); all contracts without data spanning contract start to end; IMSs in PDF or picture format; IMSs in Primavera (no active license). Currently 12 contracts with 133 schedule observations. Uniform data extraction methodology used across contract schedules: extracted 19 standard fields for non-summary activities; converted into Microsoft Excel flat-files and used standard template for metrics gathering and analysis. Additional contract insight gained from Defense Acquisition Management Information Retrieval (DAMIR).

[3] Cf. "Ending the EAC Tail Chase: An Unbiased EAC Predictor using Progress Metrics, Eric R. Druker, et al., SCEA/ISPA, 2007.

Because only the IMS itself can answer the question "when will the deliverable arrive?" a number of schedule metrics have been devised. For example, the Baseline Execution Index (BEI) indicates the efficiency with which tasks have been accomplished when measured against the baseline tasks.[4] However, the BEI merely gives the number of actually tasks finished compared to the number of tasks planned, with no regard for relative task importance, lateness, or sequencing. One cannot develop risk-adjusted end dates with this metric, but may glean task performance. Similarly, the Critical Path Length Index (CPLI), which measures the efficiency required to complete a milestone on time,[5] only gives some signal of schedule risk and is driven by expected future performance. Guidance from authoritative sources such as the GAO Schedule Assessment Guide, the DCMA 14-Point Schedule Metrics for IMS, and the Joint Cost and Schedule Risk and Uncertainty Handbook, focus primarily on gauging the quality of the most current IMS submission. This is because the IMS is a "living document" where only the latest IMS incorporates up-to-date information. However, many still find it difficult to ascertain the realism of any given schedule because current schedule quality is a necessary but not sufficient criterion for schedule realism.

The inadequacy of status quo metrics stems from the perceived continuing irrelevancy of previous IMS submissions. Without a point of reference to ensure logical evolution, the current IMS can only say so much. It is important to understand, for example, what baseline changes have occurred over time and how actual performance has measured to near-term plans. While schedules are living documents, the initial baseline stands as the only available point of reference. This baseline is valid for three major reasons. First, planners tend to know the major activities involved in the execution of a project. All systems can be said to have historical analogies, even those considered revolutionary. Second, contractors generally have well-defined processes for developing these systems. Third, the IMS undergoes an Integrated Baseline Review (IBR) after which both the contractor and client agree to the plan and its efficacy. Thus, the IMS from its outset may be viewed by estimators primarily as a tool for measuring the scheduler's ability to plan.

This paper argues that an activity's baseline from the initial IMS is relevant through subsequent submissions. Planners generally do a good job of laying out major activities, and so early performance on near-term activities should be a good indicator of total schedule realism. Practitioners have often heard anecdotally that early schedule loss cannot be made up despite managerial tactics. This paper will explore that notion and how it finds support in the data. It will be shown that by tracing near-term activities through subsequent IMSs and comparing them to their original baseline, as opposed to the current baseline, one may extrapolate a more realistic contract finish date far earlier in the project. Part I will lay the groundwork for the method's rationale by illustrating how real IMS MDAP data behaves over time. Part II will formally present the new method, as well as provide examples and show results from actual contracts.

---

[4] DCMA, "Earned Value Management System (EVM) Program Analysis Pamphlet (PAP)," pp. 16-18.
[5] DCMA, "Earned Value Management System (EVM) Program Analysis Pamphlet (PAP)," pp. 18-20.

Part 1: IMS Observations and Relationships

This section will reveal facts and trends about IMS data across the lifecycle of a contract. It will be shown that, in general, projects rely on forecasts with expectations of task performance which become increasingly detached from historical actuals. Not only are forecasts biased toward undue optimism, but schedules quality tends to decrease significantly over time leading to IMSs with relatively few interrelationships between activities. The charts that will be shown use a variety of x-axes. In normalizing schedule duration, there are three versions of % schedule used: original, current, and latest (see depiction below). The last x-axis measure used is "# Months Past Initial IMS" in which data are binned according to the number of months the IMS was submitted from the very first IMS (non-normalized).

$$\%\ \text{Schedule}_x = \frac{(\text{Schedule 'As Of'} - \text{Contract Start}_0)}{(\text{Predicted End}_x - \text{Contract Start}_0)}$$

$$\text{where Predicted End}_x = \begin{cases} \text{First IMS (Original)} \\ \text{'As Of' IMS (Current)} \\ \text{Last IMS (Latest)} \end{cases}$$

The mean actual BEI values, binned by % schedule (current) in blue, is depicted in Figure 3 below. The corresponding mean forecast BEI is in red. BEI values of 1.0 represent completing the same number of tasks as had been planned by the submission's 'as of' date. Values above 1.0 represent working ahead of plan and value below 1.0 represent falling behind in task completion. That the average forecast task performance jumps well above the actuals strikes one immediately. Anticipated efficiencies increase dramatically even though actual performance tends not to stray far from 1.0. The spread between these two metrics mirror their cost counterparts, the Cost Performance Index (CPI) and the To-Complete Cost Performance Index (TCPI). When forecasted values lay well above their historical actuals, there is significant risk as the project schedule relies on unreasonable task performance.

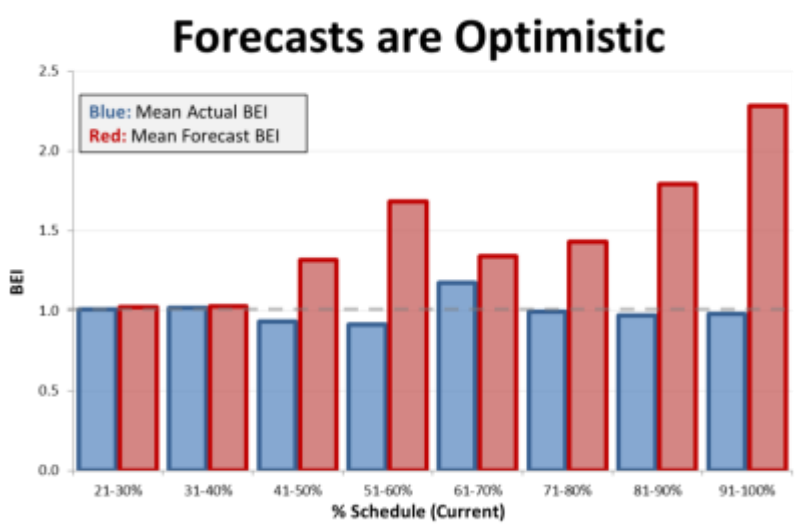


Figure 3: Actual BEI (blue) and Forecasted BEI (red). BEI = 1.0 denotes working to plan.

Though schedules rely on completing tasks on time, the BEI does not give a holistic measure of efficiency. As stated earlier, it does not measure task lateness or importance, but rather aggregate completions irrespective of sequencing. Additionally, as the project progresses, the weight of past activities starts to drive the BEI metric. For example, a project has completed 150 tasks compared to 200 planned, the BEI is 0.75. However, if 450 tasks were completed to 500 planned, the BEI is 0.90. The BEI, then, trends towards 1.0 even though the contractor remains behind at a constant rate of 50 tasks.[6]

Task performance may also be readily evaluated by counting the number of tasks which finished late relative to their baseline dates. Figure 4 below shows just that: the percentage of discrete tasks which finished late to baseline over the 9 months prior to submission 'as of' date. This sample shows task performance degradation over time. The interpretation of a linear slope is that for every 2% of schedule that passes, an additional 1% of recently finished tasks will finish late to baseline. This observation strengthens the case that future schedule optimism is unfounded. Rather than completing tasks more efficiently in the future, reality suggests that more tasks than not will finish late.

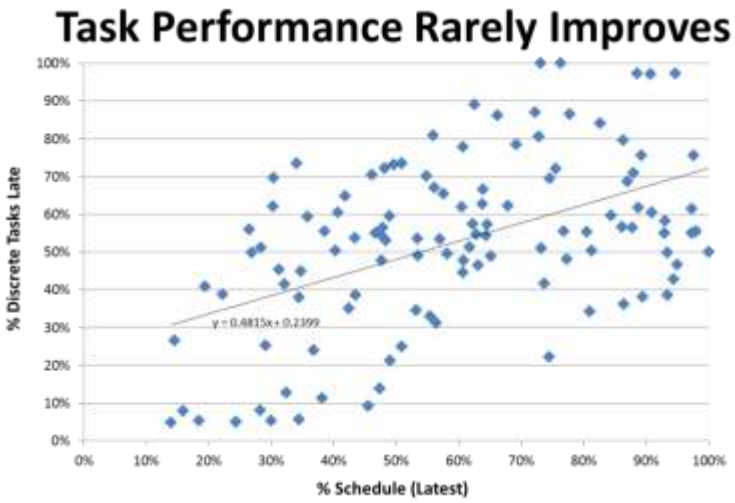


Figure 4: % Discrete Tasks which finish late to current baseline over past 9 months

Thus far we have been seeing tasks performance relative to their current baseline. But how has that baseline changed over time when compared to the very first IMS? For each change, has there been an approved Baseline Change Request (BCR)? The blue bars in Figure 5 below show the average percentage of activities with a change in baseline finish date. Of activities which persist through IMS submissions, roughly a quarter of them have changes to their baseline. On average, those activities with baseline changes eventually slip over 2 months (red scatter, right axis) and in several cases well over a year (not depicted).

[6] A similar situation is true of cost metric counterparts like cumulative Cost Performance Index (CPI) and Schedule Performance Index (SPI).

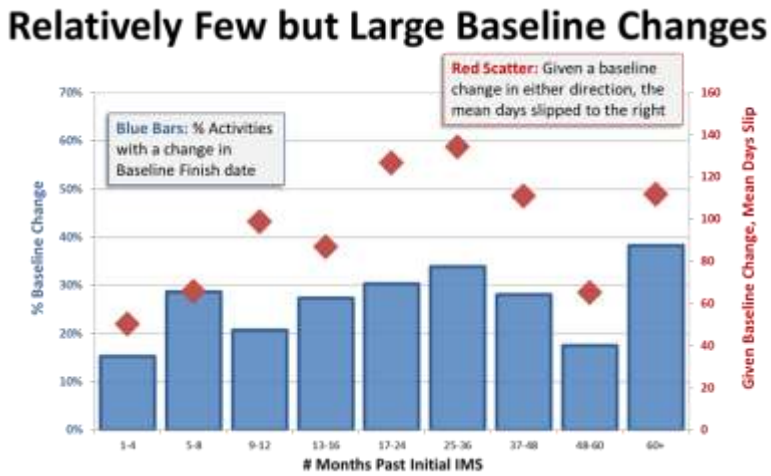


Figure 5: % Baselines change (blue, left axis). Mean days change of new baseline (red, right axis)

Aside from task performance, another key metric used to measure schedule quality is the occurrence of constraints. The DCMA 14-Point Check flags any IMS with more than 5% hard constraints as they prevent tasks from being moved by their dependencies and, therefore, prevent the schedule from being logic-driven.[7] In Figure 6 below, one may see that many schedules constrain more than 5%, and up to 20%, of activities (blue). Of those constraints, generally 50% or more are binding. Binding means that the forecast start or end date equals the constraint date. When a constraint binds an activity, it is not driven by its predecessors and may no longer slip. Though an activity may have a sound reason for requiring a hard constraint, such tasks can distort the IMS and mask a probable schedule slip.

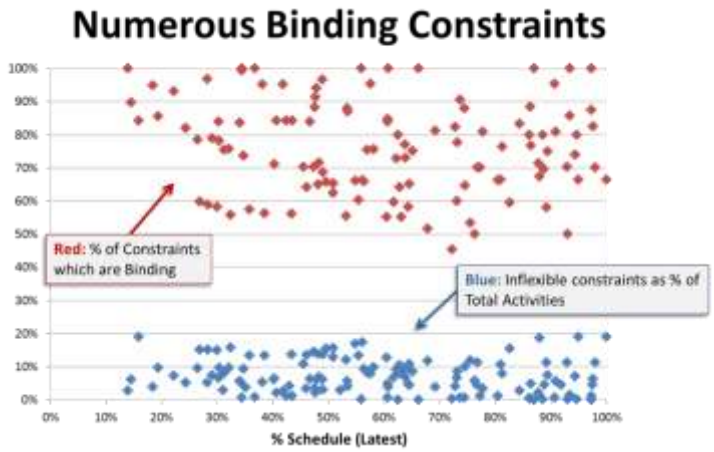


Figure 6: (Blue) % of total activities constrained. (Red) % of constrained activities which are binding.

Maybe the most important concept in schedule construction and maintenance is logic: the idea that every activity must have at least one predecessor and one successor activity. Logic drives the sequencing of activities and interconnects the schedule so that a change to any one given activity has a ripple on all successors. The DCMA 14-Point check also deems it risky for more than 5% of incomplete

[7] DCMA, "Earned Value Management System (EVM) Program Analysis Pamphlet (PAP)," pp. 16-27.

activities to have missing logic in an IMS.[8] Figure 7 below depicts the number of tasks missing logic forecasted to finish over the next three months. It is apparent that more than 5% from this subset of tasks do not have logic links. Additionally, the percentage of tasks without logic tends to grow over time with certain schedules missing up to 80%. This indicates that schedule quality tends to decrease over time as an increasing number of near-term tasks are not tied to the schedule logic.

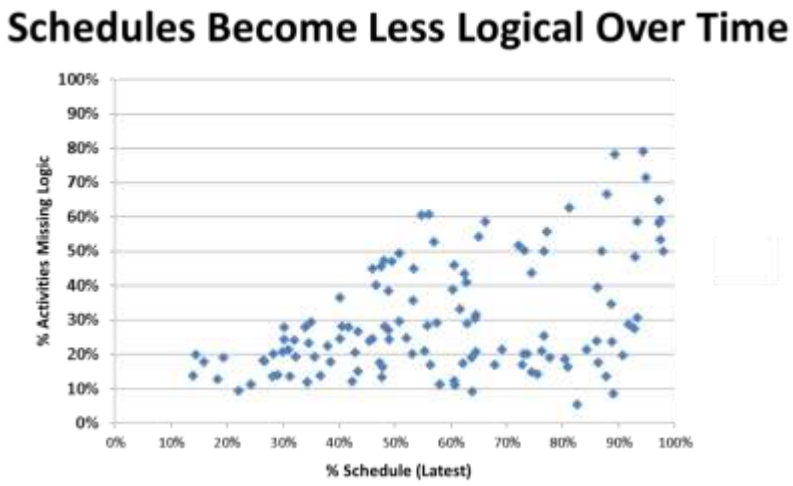


Figure 7: % of activities missing both predecessor & successor logic links in next 3 months

The preceding section has shown that, in general, schedule quality and performance tend to decrease over the course of a contract. A high number of binding constraints and increasing number of late tasks leads to schedule optimism. Baseline changes without BCRs can conceal poor performance. Fewer logic links indicate a poorly maintained schedule. These generalizations have a negative impact on schedule realism. Even a high quality rating for any given IMS does not mean that it has also evolved in a consistent manner. The relevant question becomes "can current IMS data be better utilized to measure schedule risk and extrapolate a realistic end date far earlier in the project?" Part II of this paper will explore a process for doing just that.

[8] DCMA, "Earned Value Management System (EVM) Program Analysis Pamphlet (PAP)," pp. 25.

Part 2: An Improved Schedule Estimator

The lynchpin behind the proposed schedule estimating technique is the baseline IMS. The first available IMS, preferably the one immediately following IBR, will be used to independently track activities over subsequent submissions. Activity end dates are traced over time using their designated activity names and/or unique IDs. These updated end dates will then be compared to their original baseline end date and total float (slack days).[9] This process disregards all: new activities which get built into the schedule over time; activities which change identifiers; and new activity sequencing. While seemingly concerning, remember, the primary purpose is to trace performance of the relatively near-term plan. And both contractual parties agreed to the baseline IMS which took months to develop. That some activities cannot be traced (e.g. a task broken up into three new, smaller tasks) is simply a reality insofar as there is no record of the changes.

In order to determine schedule slip, one looks across all tasks still identifiable in successive IMS submissions and compares them to their original baseline. The activity which has slipped most to that baseline, with the original days of float factored in, drives the independently predicted end date. That number of days slip is added to the original contract closeout date.

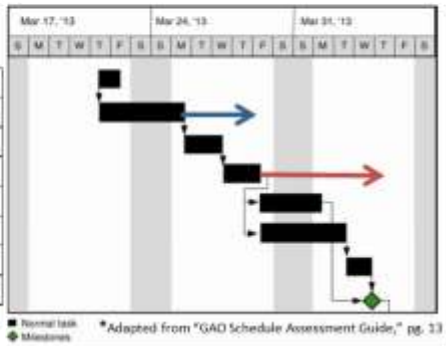
$$\text{Predicted End Date}(t_n) = \text{Baseline End Date}(t_0) + \text{Slip}(t_n)$$
$$\text{Slip}(t_n) = \max(\{f(x_i): i = 1,\ldots,k\})$$
$$\text{where } f(x_i) = [\text{Current Finish}_i - \text{Baseline Finish}_i - \text{Baseline Total Float}_i]$$
$$\text{and } \max(\{f(x_i): i = 1,\ldots,k\}) \geq 0$$
$$x_i = \text{observed activity}$$

When the schedule has evolved to the extent that the current activities no longer reflect the baseline, the predicted end date stabilizes and the estimate is determined. This occurs because the tasks in the baseline IMS have their finish dates realized and the planning packages open up into new work packages. The rule implemented by the author is to stop the analysis when more than 95% of the activities from the original baseline are either finished or no longer appear in the current schedule. The final prediction for contract end is then set. An example will illustrate this methodology.

[9] Total float (also called total slack) is the number of days an activity is allowed to move to the right before any further slips affect the project end date. When an activity has 0 days of slack, it might find itself on the "critical path," or a sequence of activities representing the shortest duration between the schedule's 'as of' date and the end of the project.

An Example

	IMS #1		IMS #2		IMS #3	
Activity Name	Baseline Finish	Total Float	Finish	Slip	Finish	Slip
Rough grade property	03-20-13	3	03-20-13	0	03-20-13	0
Frame first floor walls	03-25-13	3	03-29-13	1	03-29-13	1
Install roof trusses	03-27-13	3	03-29-13	0	03-29-13	0
Install roof decking	03-29-13	3	03-31-13	0	04-05-13	4
Rough-in framing inspection	04-01-13	5	04-03-13	0	04-03-13	0
Form and pour driveway	04-02-13	3	04-04-13	0	04-05-13	0
Finish excavate and pour garage	04-03-13	3	04-05-13	0	04-06-13	0
Framing complete	04-03-13	3	04-05-13	0	04-06-13	0
			Max Slip:	1	Max Slip:	4



*Adapted from "GAO Schedule Assessment Guide," pg. 13

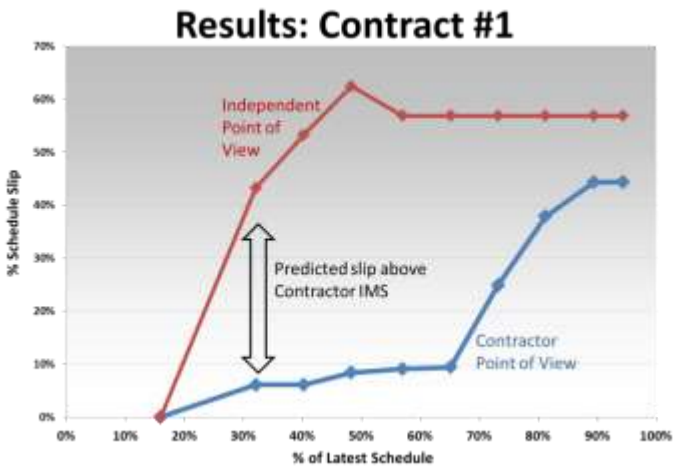
Figure 8: A simplified example of an evolving IMS.

The first IMS sets the baseline for future activities. All activity finish dates will be evaluated for each subsequent IMS and compared to their baseline. In the case of IMS #2, several activities have experienced slips in their forecast finish dates. However, "Frame first floor walls" has slipped 4 days while it only had 3 days of float available. This means that the task will affect the start of all of its successor tasks and push out the milestone "Framing complete" by 1 day, *ceteris paribus*. In IMS #3, "Install roof decking" has slipped 7 days to baseline, and factoring in its original 3 days of float, should now affect "Framing Complete" by 4 days.

How has "Framing complete" been able to stay on track? In the example above, the forecasted tasks have had their durations implicitly squeezed in order to make up time. In reality, forecasted discrete tasks do not regularly show shorter durations than past tasks. A plausible reason is that the long-term future is represented by planning packages which incorporate many tasks. When opened, the resulting tasks are given realistic durations and distribute the implicit duration squeeze onto new planning packages. Although evidence cannot be concretely pulled from the data, it has been shown that forecasted performance is optimistic and there are numerous binding constraints. The occurrence of these behaviors is sufficient to maintain a fixed near-term schedule. The total schedule might be saved, however, because of the future's vagueness. As time progresses, the pool of undetailed tasks dwindles and increasingly focuses implicit duration compression. By the time the schedule's unrealism is noted (possibly by pure reckoning), it would be relatively late in the project.

The proposed technique disregards over-optimism by relying on early task performance to baseline. Because major activities are reasonably phased from the start, the total schedule receives a one-for-one slip with the near-term (realized) schedule. The idea becomes increasingly attractive when considering the fact that schedules tend to lose an alarming number of logic links over time. This means tracking near-term performance late in the schedule tells one little, since relatively few tasks are imparted with interrelationships. The estimation technique is not intended to project exact dates, but

should be viewed as a tool to measure the magnitude of a would-be schedule slip. The results of this technique are presented for 8 MDAP contracts in a standard chart template.[10]



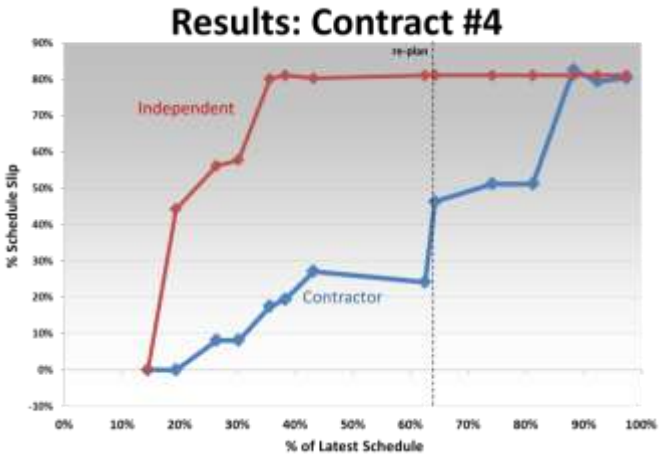
Above, one may see that from the contractor's point of view, the IMS shows less than 10% predicted schedule slip up until near 70% of the latest schedule. Then significant delays are realized, and the project eventually slips by almost 45% to baseline. The independent point of view, representing the technique proposed in this paper, quickly registers schedule risk to within about 10% of actual slip. Note that the (red) independent line plateaus shortly after 50% of latest schedule. This is where the current schedule largely reflects new activities; while those from the baseline IMS have for the most part been completed.

Results from a second contract are similar to the first: the independent technique gives a good approximation of the realized schedule slip fairly early on in the contract. This contract, however, has a

[10] Though 12 MDAP contracts were collected and analyzed, 4 were of the type Indefinite Delivery, Indefinite Quantity (IDIQ). The 4 IDIQ contracts are not shown because new contract task orders or scope changes have a clear effect on schedule duration. More data insight is needed to compensate for the volatility in scope.

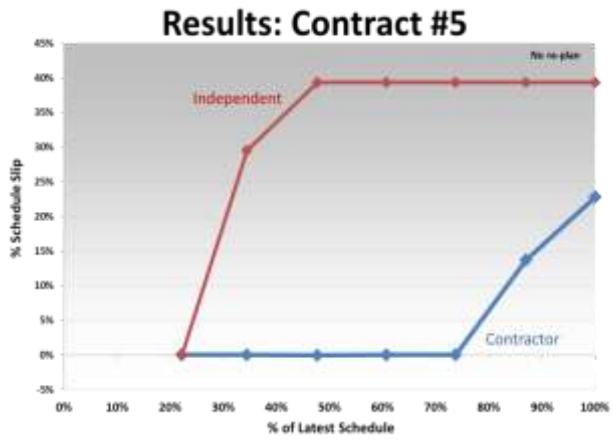
major re-plan shortly after 40% of latest schedule.[11] Because the independent line plateaus first, it has the expected duration of the slip associated with the re-plan embedded in it. This is valid because the project did not increase quantity or add significant capabilities. Additional budget and schedule were allocated to a struggling project. This method is useful where re-plans are an admission of *de facto* poor performance to plan because of technical problems or otherwise. This method does not do a good job in its raw form presented here for Indefinite Delivery, Indefinite Quantity (IDIQ) contract types.

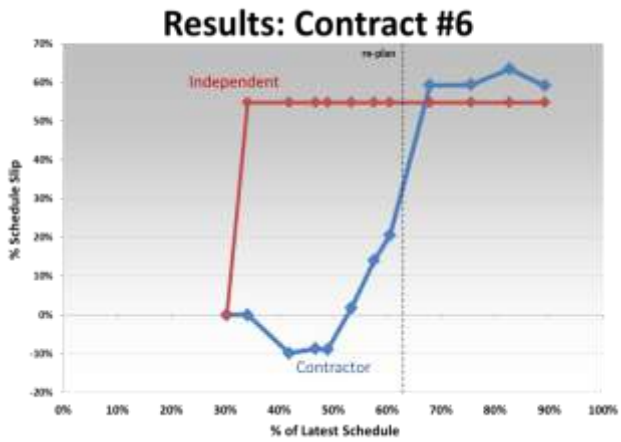
The contract above shows a situation where there was a re-scope early in the contract. Because a new plan was implemented before the independent metric was able to measure the realized slip in near-term baseline tasks, it was unable to register the extent of the total schedule slip. In such situations, the new IMS may be used as a fresh baseline, and the analysis restarts. The contract above underwent a second major re-plan shortly thereafter.

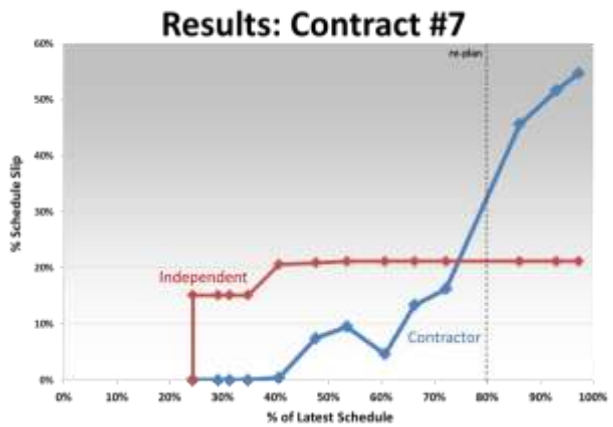


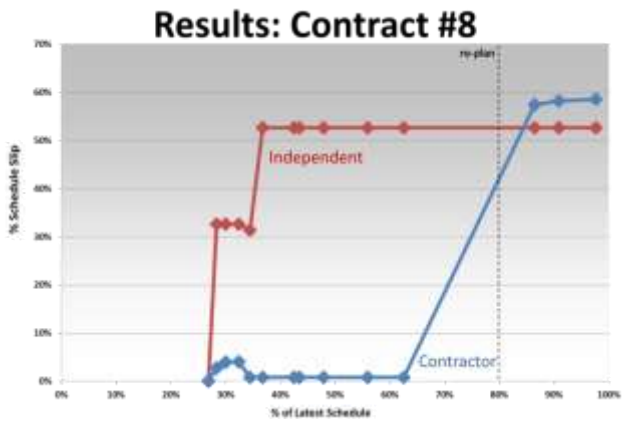
[11] Re-plans are ascertained by inspecting the Contract Performance Reports (CPRs) Formats 1-5 as well as contract narratives from Defense Acquisition Management Information Retrieval (DAMIR).

Again, one may see that the independent metric is close to measuring the true schedule slip early in the contract. In the results shown above, the technique quickly shows 40% then 80% slip. Though the project eventually conformed to that prediction, it brings up the problem of the self-fulfilling prophecy. If a decision-maker admits to significant delays early on, then the project will be managed to a longer timeframe and the personnel won't feel schedule pressure. This could, in fact, lead to further delays down the line.









The independent metric finds a useful application in cost estimation as well. If the metric predicts a one year slip over that reported in the IMS, one may extend the average project cost burn rate out for the same duration. Figure 9 below does just that using one of the real contracts analyzed in this paper. The red line is the mean actual burn rate multiplied by the number of months slip the independent metric predicts over the IMS plus the cost of the traditional independent estimate at complete (IEAC). The new independent cost estimate acts similar to the schedule estimator and gives a good early indication of the realized costs.

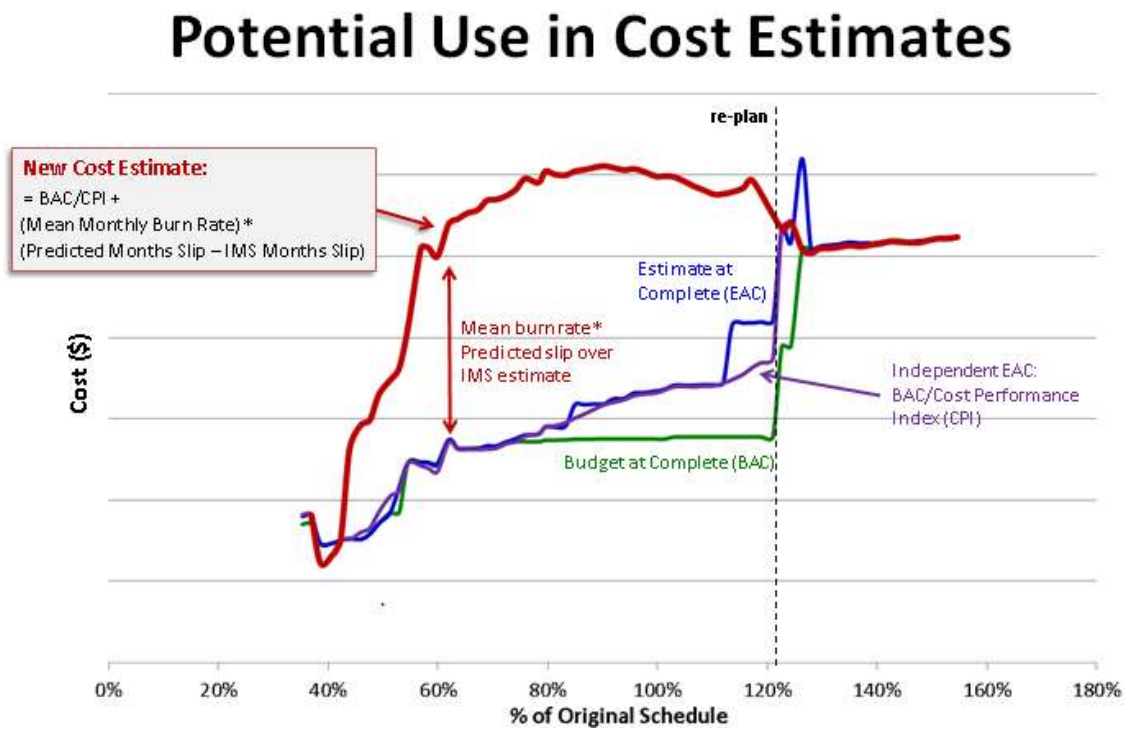


Figure 9: Application of independent metric to cost estimation for an actual contract

How may an estimator use this metric to improve outcomes on MDAP contracts? First and foremost, it gives the decision-maker an early indicator to the magnitude of a potential schedule slip. Figure 10 below illustrates the new metric's predictive power. By 50% of the latest schedule duration, the independent metric estimates a majority of realized schedule slip. The mean absolute error for the 41-50% bin is only 7 months, compared the contractor's 25 months. This allows management to plan early for cost-schedule-technical tradeoffs. Early recognition of catastrophic difficulties also leads to greater flexibility in project termination because of fewer costs sunk. It may also prove useful in providing more realistic cost estimates. Additionally, the method allows for traceability between submissions. Because high schedule quality as traditionally calculated may be misleading, it is useful to understand how volatile activities are and whether the IMS is continually being re-invented.

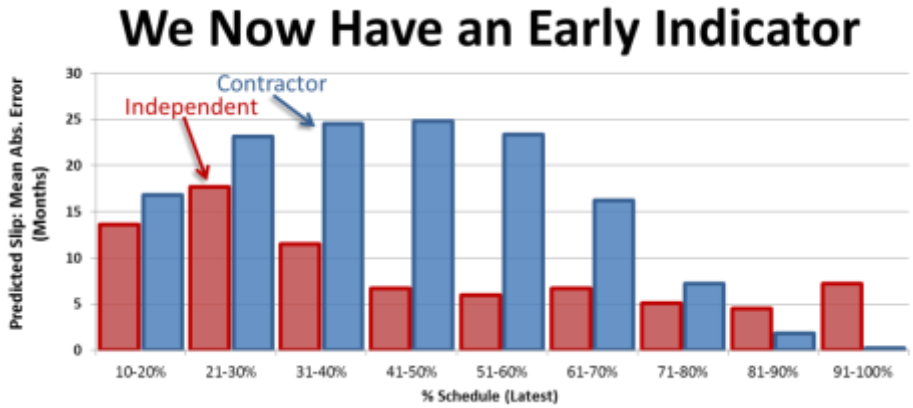


Figure 10: Mean absolute error of contractor and independent schedule estimates

Using this new technique, one may catch potential schedule slip early in the project with relative accuracy. While schedules develop over time, planners do a good job at the outset of phasing major milestones and near-term work packages. Though projects change, the deliverable doesn't often transform completely. However, cumulative changes to a schedule tend to deteriorate its realism. As we have seen, schedule performance is best registered early for a variety of reasons including optimistic forecasts and decreasing quality. One may conclude that contractors quickly reveal their pace of work and "settle" into a performance; or management strategies such as work-arounds can do little to alleviate ailing projects. Until a culture of improved schedule maintenance takes root, performance to baseline early in the project may serve as a good indication of realized schedule slip.

There is much to be done using this schedule estimating technique in future refinements. First and foremost is the need for an expanded data set to repeatedly test the metric's robustness. It must also follow a more detailed approach, such as analysis by task-order, to assess the value for IDIQ contracts. Further assessment is also needed for augmentation to cost estimation. One method (extending mean burn rates) has been presented in this paper, but many other approaches are possible. Finally, a large collection of IMSs used for this analysis can also provide data-driven generalizations of schedule quality and realism over time. Additional insights into IMS evolution can help practitioners understand how to better manage their schedule.

Bibliography:

"Earned Value Management System (EVMS) Program Analysis Pamphlet (PAP)." Defense Contract Management Agency (DCMA). July 2012. http://www.dcma.mil/policy/200-1/PAM-200-1.pdf. Accessed 10 April 2014.

"Ending the EAC Tail Chase: An Unbiased EAC Predictor using Progress Metrics, Eric R. Druker, et al., SCEA/ISPA, 2007.

"GAO Schedule Assessment Guide." U.S. Government Accountability Office (GAO). 30 May 2012. http://www.gao.gov/assets/600/591240.pdf. Accessed 10 April 2014.

"IMS DID." Office of the Undersecretary for Defense Acquisition, Technology, & Logistics Performance Assessments and Root Cause Analyses (OUSD AT&L PARCA). 30 March 2005. http://dcarc.cape.osd.mil/EVM/Documents.aspx. Accessed 10 April 2014.

"IPMR Implementation Guide." Office of the Undersecretary for Defense Acquisition, Technology, & Logistics Performance Assessments and Root Cause Analyses (OUSD AT&L PARCA). July 21 2012. http://dcarc.cape.osd.mil/EVM/Documents.aspx. Accessed 10 April 2014.

"Joint Cost and Schedule Risk and Uncertainty Handbook (CSRUH)." Naval Center for Cost Analysis (NCCA). 16 July 2013. https://www.ncca.navy.mil/tools/csruh/index.cfm%20. Accessed 10 April 2014.